

# On the primal-dual dynamics of Support Vector Machines

Krishna Chaitanya Kosaraju, Shraavan Mohan and Ramkrishna Pasumarthy

**Abstract**—The aim of this paper is to study the convergence of the primal-dual dynamics pertaining to Support Vector Machines (SVM). The optimization routine, used for determining an SVM for classification, is first formulated as a dynamical system. The dynamical system is constructed such that its equilibrium point is the solution to the SVM optimization problem. It is then shown, using passivity theory, that the dynamical system is global asymptotically stable. In other words, the dynamical system converges onto the optimal solution asymptotically, irrespective of the initial condition. Simulations and computations are provided for corroboration.

## I. INTRODUCTION

The field of Machine Learning has gained tremendous traction over the past decade with the advent of data compilation from various sectors of the industrial world [1]. The techniques therein have helped the industry gain crucial insights into their processes and make judicious decisions for the future. A ubiquitous component of most Machine Learning algorithms is optimization, where in a suitably chosen cost function is maximized (or minimized) under constraints. In many applications, the cost function and the constraints arise from practical considerations. As far as the optimization routines are concerned, the most well understood class of optimization problems happens to be that of convex optimization [2]. Convex optimization problems happen to be quite useful and have also percolated into many different application areas. A particularly interesting application is that of classification problems using Support Vector Machines. Support vector machines form a tool set for linear as well as non-linear classification [3]. They can also be used effectively for non-linear regression using different kernels. As such, classification itself turns out to be quite useful in the industry; applications range from predicting defaulters in finance sector, predicting claims in the insurance sector and detecting defects in retinopathy [4], [5].

Gradient-based methods form a fundamental basis of all algorithms for solving convex optimization problems. These gradient algorithms have much to gain from a control and dynamical systems perspective; especially for a better understanding of the underlying system theoretic properties such as stability, convergence rates, and robustness [6]. The

convergence of gradient-based methods and Lyapunov stability, relate the solution of the optimization problem to the equilibrium point of a dynamical system [7], [8], [9], [10]. In this context, the focus of this paper is continuous time primal-dual gradient descent method. The formulation has its roots from [11], where the author constructs a dynamical system whose trajectories converges asymptotically to the solution of a min-max problem (saddle point problem). This framework has two very important characteristics. The first being the equilibrium of the dynamical system is not explicitly known but it is implicitly characterized by the Karush-Kuhn-Tucker (KKT) conditions of an optimization problem (or the optimization problem itself). Secondly, the fact that one can show stability using Lyapunov analysis without the knowledge of equilibrium set or a point. In the literature, such systems are usually called *contracting systems*, a term coined in the seminal paper [12], where the authors show that the distance between the trajectories contracts exponentially.

*Motivation and contribution:* In this paper, we consider a convex optimization formulation of a linear support vector machine problem (usually noted as primal formulation). We next propose the Lagrangian of the constrained optimization problem using which we present its dual formulation. The primal together with its dual forms gives rise to a saddle-point problem. We present the continuous time gradient descent equation for the saddle-point problem, which essentially captures two properties; minimization of the Lagrangian with respect to the primal variables and maximization of Lagrangian with respect to the dual variables [9], [10]. Hence these dynamics are usually noted as primal-dual dynamics. We finally present the convergence analysis of these dynamics using tools from passivity [13] and hybrid systems theory [14]. Note that, rewriting the algorithm as dynamical system that converges to the solution of an optimization problem has enabled the use of such systems theory tools for convergence analysis. The main objective of this note is to motivate the dynamical system formulation which will acts as a fundamental entity for future research. Simulation studies are provided to understand the behavior of Lyapunov function and visualizing the results.

The paper is organized as follows. Section II presents a brief overview of convex optimization. Section III presents results on the stability of the continuous time primal-dual dynamics used to solve convex optimization problems. Finally, in Section IV, the ideas are applied to the case of the SVM and simulations are provided for corroboration.

Krishna Chaitanya and Shraavan Mohan are graduate students in the Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai-600036, India. kkrishnachaitanya89@gmail.com, shraavan8587@gmail.com

Ramkrishna is with the Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai-600036, India ramkrishna@iitm.ac.in

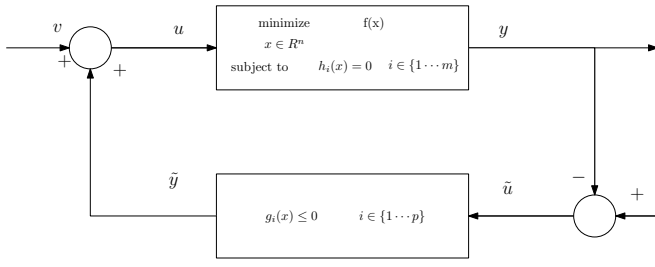


Fig. 1. Interconnected optimization

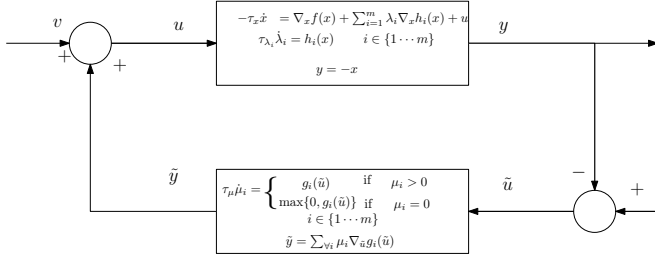


Fig. 2. Interconnected primal dual dynamics

## II. CONVEX OPTIMIZATION

In this section, we present a brief overview of mathematical tools in convex optimization, that will be useful in the subsequent sections. The standard form of a convex optimization problem contains three parts:

- (i) A continuously differentiable convex function  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$  to be minimized over  $x$ ,
- (ii) affine equality constraints  $h_i(x) = 0, \quad i = 1, \dots, m$ ,
- (iii) continuously differentiable convex inequality constraints of the form  $g_i(x) \leq 0, \quad i = 1, \dots, p$ .

This can be written in the following form, commonly known as the primal formulation:

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && f(x) \\ & \text{subject to} && h_i(x) = 0 \quad i = 1, \dots, m \\ & && g_i(x) \leq 0 \quad i = 1, \dots, p \end{aligned} \quad (1)$$

**Karush-Kuhn-Tucker (KKT) conditions:** If the solution  $x^*$  is optimal to the convex optimization problem (1) then there exist  $\lambda_i \in \mathbb{R}, i = 1, \dots, m$  and  $\mu_i \geq 0, i = 1, \dots, p$  satisfying the following KKT conditions

$$\begin{aligned} \nabla_x f(x^*) + \sum_{i=1}^m \lambda_i \nabla_x h_i(x^*) + \sum_{i=1}^p \mu_i \nabla_x g_i(x^*) &= 0, \\ h_i(x^*) &= 0 \quad \forall i \in \{1, \dots, m\}, \\ g_j(x^*) &\leq 0, \quad \mu_j \geq 0, \quad \mu_j g_j(x^*) = 0 \quad \forall j \in \{1, \dots, p\}. \end{aligned} \quad (2)$$

**Remark 1.** Note that the KKT conditions presented above in equation (2) are only necessary conditions. We next present the requirements under which KKT conditions become sufficient.

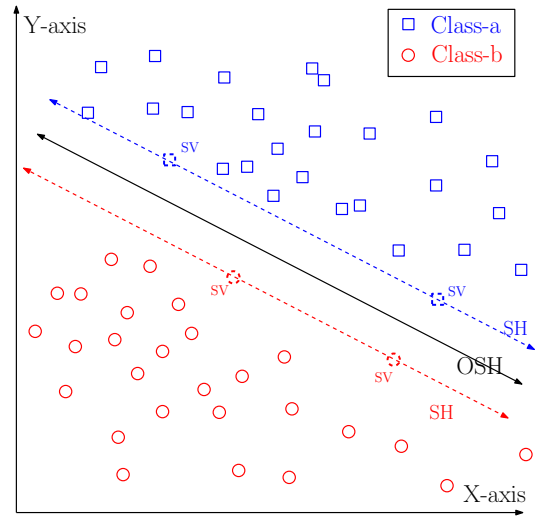


Fig. 3. Description of a linear support vector machine.

We now define the Lagrangian of the convex optimization (1) as

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{i=1}^p \mu_i g_i(x) \quad (3)$$

and the Lagrange dual function as

$$L_d(\lambda, \mu) = \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \mathcal{L}(x, \lambda, \mu) \quad (4)$$

giving us the following dual problem (corresponding to the primal problem (1))

$$\begin{aligned} & \underset{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p}{\text{maximize}} && L_d(\lambda, \mu) \\ & \text{subject to} && \mu_i \geq 0 \quad i = 1, \dots, p. \end{aligned} \quad (5)$$

**Remark 2.** Dual problem is always convex, because  $L_d$  is always a concave function even when the primal (1) is not convex. If  $f^*$  and  $L_d^*$  denotes the optimal values of primal and dual problems respectively, then  $L_d^* \leq f^*$ . Therefore dual formulations are used to find the best lower bound of the optimization problem [2], [15]. Further, the negative number  $L_d^* - f^*$  denotes the *duality gap*. In the case of zero duality gap, we say that the problem (1) *satisfies strong duality*.

**Definition 1. Slater's conditions.** The convex optimization problem (1) is said to satisfy Slater's conditions if there exists an  $x$  such that  $h_i(x) = 0 \quad i = 1, \dots, m$  and  $g_i(x) < 0 \quad i = 1, \dots, p$ . This implies that inequality constraints are strictly feasible.

**Remark 3.** If a convex optimization problem (1) satisfies Slater's conditions then the optimal values of primal and dual problems are equal, that is, (1) satisfies strong duality. Further, in this case the KKT conditions becomes necessary and sufficient.

## III. STABILITY OF PRIMAL-DUAL DYNAMICS

In this section, we present the continuous time primal-dual equations of a convex optimization problem. In [10], we have

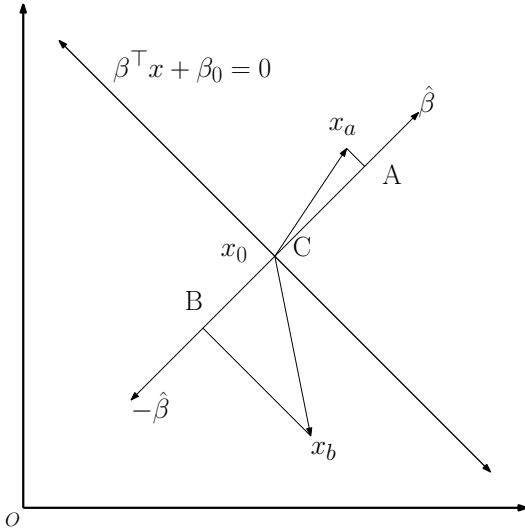


Fig. 4. Mathematical formulation of a linear support vector machine,  $x_a \in X_a$  (class-a) and  $x_b \in X_b$  (class-b).

shown that these dynamics can be described as a feed-back interconnection of two passive dynamical systems. The first being the primal-dual dynamics of an equality constrained optimization problem and the second corresponds to the hybrid dynamics representing the inequality constraint (see Figures 1, 2). We now briefly revisit these results.

Assume that Slater's condition holds. Since strong duality holds for (1),  $(x^*, \lambda^*, \mu^*)$  satisfying the KKT conditions (2) is a saddle point of the Lagrangian  $\mathcal{L}$ . This implies,  $x^*$  is an optimal solution to primal problem (1) and  $(\lambda^*, \mu^*)$  is optimal solution to its dual problem (5), that is

$$(x^*, \lambda^*, \mu^*) = \arg \max_{\lambda, \mu} \left( \arg \min_x \mathcal{L}(x, \lambda, \mu) \right). \quad (6)$$

This gives us the following saddle-point dynamics.

$$\begin{aligned} -\tau_x \dot{x} &= \left( \nabla_x f(x) + \sum_{i=1}^m \lambda_i \nabla_x h_i(x) + \sum_{i=1}^p \mu_i \nabla_x g_i(x) \right) \\ \tau_\lambda \dot{\lambda} &= h(x) \\ \tau_\mu \dot{\mu} &= (g(x))_\mu^+ \end{aligned} \quad (7)$$

$\tau_x, \tau_\lambda \triangleq \text{diag}\{\tau_{\lambda_1}, \dots, \tau_{\lambda_m}\}$ ,  $\tau_\mu \triangleq \text{diag}\{\tau_{\mu_1}, \dots, \tau_{\mu_p}\}$  are positive definite matrices and  $(g(x))_\mu^+$  is given by

$$(g(x))_\mu^+ = \begin{cases} g_i(x) & \text{if } \mu_i > 0 \quad \forall i \in \{1, \dots, p\}, \\ \max(0, g_i(x)) & \text{if } \mu_i = 0. \end{cases} \quad (8)$$

**Remark 4.** The discrete time primal-dual gradient descent equations of convex optimization problem (1) are

$$\begin{aligned} x(t_{k+1}) &= x(t_k) - \eta_x \nabla_x \mathcal{L}(x, \lambda, \mu) \\ \lambda(t_{k+1}) &= \lambda(t_k) + \eta_\lambda \nabla_\lambda \mathcal{L}(x, \lambda, \mu), \quad k \in \mathbb{Z}^+ \\ \mu(t_{k+1}) &= \mu(t_k) + \eta_\mu (\nabla_\mu \mathcal{L}(x, \lambda, \mu))_\mu^+, \quad k \in \mathbb{Z}^+. \end{aligned}$$

where  $\eta_x > 0$ ,  $\eta_\lambda > 0$  and  $\eta_\mu > 0$  represents the step size. Further these are equivalent to the continuous time equations

(7), if the step sizes are chosen as  $\eta_x = \Delta T \tau_x^{-1}$ ,  $\eta_\lambda = \Delta T \tau_\lambda^{-1}$  and  $\eta_\mu = \Delta T \tau_\mu^{-1}$ , where  $\Delta T = t_{k+1} - t_k$ .

**Equality constrained optimization problem:** Consider the following dynamics

$$\begin{aligned} -\tau_x \dot{x} &= \nabla_x f(x) + \sum_{i=1}^m \lambda_i \nabla_x h_i(x) + u \\ \tau_{\lambda_i} \dot{\lambda}_i &= h_i(x), \quad y = -x, \end{aligned} \quad (9)$$

where  $u, y \in \mathbb{R}^n$ . Note that, the unforced system of equations, obtained by setting  $u = 0$  in (9), represent primal-dual dynamics corresponding to convex optimization problem (1) with only equality constraints.

**Proposition 1.** Let  $\bar{z} = (\bar{x}, \bar{\lambda})$  represent the unforced equilibrium of (9). Assume  $h(x)$  is convex and  $f(x)$  strictly convex. Then the system of equations (9) are passive with port variables  $(\dot{u}, \dot{y})$  [10], [16]. Further every solution of the unforced version ( $u = 0$ ) of (9) asymptotically converges to  $\bar{z}$ .

**Inequality constraint:** We now define the inequality constraint  $g_i(\tilde{u}) \leq 0$  as the following hybrid dynamics

$$\tau_{\mu_i} \dot{\mu}_i = (g_i(\tilde{u}))_{\mu_i}^+ \quad (10)$$

where  $\tilde{u} \in \mathbb{R}^n$  and  $i \in \{1 \dots p\}$ . This is introduced in [11], where the authors construct a dynamical system which converges to the stationary solution of saddle value problems. These equations are proposed in such a way that, if the initial condition of  $\mu(t)$  is non-negative, then the trajectories  $\mu(t)$  always stay inside positive orthant  $\mathbb{R}^+$ . Note that the discontinuity in the above equations (8) occurs when  $g_i(\tilde{u}) < 0$  and  $\mu_i = 0$ , the value of  $(g_i(\tilde{u}))_{\mu_i}^+$  switches from  $g_i(\tilde{u})$  to 0. This ensures that the  $\mu_i$ 's does not go below zero. To make this more visible, we redefine these equations equivalently as follows; Let  $\mathcal{P}$  represent the power set of  $\{1 \dots p\}$ , then we define the function  $\sigma : [0, \infty) \rightarrow \mathcal{P}$  as follows

$$\sigma(t) = \{i \mid \mu_i(t) = 0 \text{ and } g_i(\tilde{u}) \leq 0 \forall i \in \{1, \dots, p\}\}. \quad (11)$$

With  $\sigma(t)$  representing the switching signal, equation (10) now takes the form of a switched system

$$\tau_{\mu_i} \dot{\mu}_i = g_i(\tilde{u}, \sigma) = \begin{cases} g_i(\tilde{u}); & i \notin \sigma(t) \\ 0; & i \in \sigma(t) \end{cases} \quad (12)$$

The overall dynamics of the  $p$  inequality constraints  $g_i(\tilde{u}) \leq 0 \forall i \in \{1 \dots p\}$  can be written in a compact form as:

$$\tau_\mu \dot{\mu} = g(\tilde{u}, \sigma) \quad (13)$$

where  $\mu_i$  and  $g_i(\tilde{u}, \sigma)$  are  $i^{\text{th}}$  components of  $\mu$  and  $g(\tilde{u}, \sigma)$  respectively. Consider the following storage function(s)

$$S_{\sigma_q}(\mu) = \frac{1}{2} \sum_{i \notin \sigma_q} \mu_i^2 \tau_{\mu_i} \quad \forall \sigma_q \in \mathcal{P} \quad (14)$$

**Proposition 2.** [10], [17] The switched system (13) is passive with switched storage functions  $S_{\sigma_q}$  (defined one for each switching state  $\sigma_q \in \mathcal{P}$ ), input port  $u_s = \dot{\tilde{u}}$  and output port  $y_s = \dot{\tilde{y}}$  where  $\tilde{y} = \sum_{\forall i} \mu_i \nabla_{\tilde{u}} g_i(\tilde{u})$ . That is, for each

$\sigma_p \in \mathcal{P}$  with the property that for every pair of switching times  $(t_i, t_j)$ ,  $i < j$  such that  $\sigma(t_i) = \sigma(t_j) = \sigma_p \in \mathcal{P}$  and  $\sigma(t_k) \neq \sigma_p$  for  $t_i < t_k < t_j$ , we have

$$S_{\sigma_p}(\mu(t_j)) - S_{\sigma_p}(\mu(t_i)) \leq \int_{t_i}^{t_j} u_s^\top y_s dt. \quad (15)$$

**Proposition 3.** *The equilibrium set  $\Omega_e$  defined by constant control input  $\tilde{u} = \tilde{u}^*$  of (10)*

$$\Omega_e = \{(\bar{\mu}, \tilde{u}^*) \mid g_i(\tilde{u}^*) \leq 0, \bar{\mu}_i g_i(\tilde{u}^*) = 0 \quad \forall i \in \{1, \dots, p\}\}$$

is asymptotically stable.

*The overall optimization problem:* We now define a power conserving interconnection between passive systems associated with optimization problem with an equality constraint (9) and an inequality constraint (10) (see Fig. 1).

**Proposition 4.** [10] *Consider the interconnection of passive systems (9) and (10), via the following interconnection constraints  $u = \tilde{y} + v$  and  $\tilde{u} = -y + \tilde{v}$ ,  $v \in \mathbb{R}^p$ ,  $\tilde{v} \in \mathbb{R}^n$ . For  $\tilde{v} = 0$ , the interconnected system is then passive with port variables  $\dot{v}$ ,  $-\dot{x}$  (see Fig. 2). Moreover for  $v = 0$  and  $\tilde{v} = 0$  the interconnected system represents the primal-dual gradient dynamics of the optimization problem (1) and the trajectories converge asymptotically to the optimal solution of (1).*

In the next section, we demonstrate the continuous-time primal-dual algorithm, on the convex optimization formulation of Support Vector Machines (SVM) technique [18].

#### IV. LINEAR SVM AS AN APPLICATION

Support Vector Machines [18] are a class of supervised machine learning algorithms which are commonly used for data classification. In this methodology, each data item is a point in  $n$ -dimensional space that is mapped to a category (or a class). Here the aim is to find an optimal separating hyperplane (OSH) which separates both the classes and maximizes the distance to the closest point from either class (as shown in Figure 3). These closest points are usually called support vectors (SV). The lines passing through support vectors and parallel to the optimal separating hyperplane are called supporting hyperplanes (SH).

*Problem formulation:* Consider two linearly separable classes, where each class (say class- $a$ , class- $b$ ) contains a set of  $N$  unique data points in  $\mathbb{R}^2$ . Let  $X_a$  and  $X_b$  denote the set of points in class- $a$  and class- $b$  respectively. In this methodology we find a hyperplane that separates the classes while maximizing the distance to the closest point from either class. Let  $L$  be an affine set that characterizes such a hyperplane, defined as follows

$$L = \{x \in \mathbb{R}^2 \mid x^\top \beta + \beta_0 = 0\} \quad (16)$$

where  $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$  and  $\beta_0 \in \mathbb{R}$ . Define the map  $l : \mathbb{R}^2 \rightarrow \mathbb{R}$  by  $l(x) = x^\top \beta + \beta_0$ . Note the following, for any  $x_0 \in L$ ,  $l(x_0) = 0 \implies x_0^\top \beta = -\beta_0$ . This implies  $l(x)$  can be rewritten as  $l(x) = \beta^\top (x - x_0)$ , which further implies

the unit vector  $\hat{\beta} = \frac{\beta}{\|\beta\|}$  is orthogonal to the line defined by the set  $L$ , that is,  $x^\top \beta + \beta_0 = 0 \iff (x - x_0)^\top \beta = 0$ .

The distance between the point  $x_a \in X_a$  and line  $L$  is  $|AC| = (x_a - x_0)^\top \hat{\beta}$  (see Fig. 4). Similarly the distance between the point  $x_b \in X_b$  and line  $L$  is  $|BC| = (x_b - x_0)^\top (-\hat{\beta})$ . We want to find an optimal separating hyperplane that is at least  $M$  units away from all the points. This implies

$$\begin{aligned} \forall x_a \in X_a, \quad (x_a - x_0)^\top \hat{\beta} &\geq M, \\ \forall x_b \in X_b, \quad -(x_b - x_0)^\top \hat{\beta} &\geq M. \end{aligned} \quad (17)$$

Define  $X \triangleq X_a \cup X_b$ , and  $Y \triangleq Y_a \cup Y_b$  where  $Y_a = \{1, \dots, 1\}$  and  $Y_b = \{-1, \dots, -1\}$ . The inequality constraints (17) can be rewritten as

$$\frac{1}{\|\beta\|} y_i (\beta^\top x_i + \beta_0) \geq M \quad (18)$$

where  $y_i = 1$  if  $x_i \in X_a$  (class- $a$ ),  $y_i = -1$  if  $x_i \in X_b$  (class- $b$ ). Finally, finding the optimal separating hyperplane can be proposed as the following optimization problem,

$$\begin{aligned} &\text{maximize}_{\beta, \beta_0} \quad M \\ &\text{subject to} \quad \frac{1}{\|\beta\|} y_i (\beta^\top x_i + \beta_0) \geq M, \quad \forall x_i \in X, y_i \in Y. \end{aligned} \quad (19)$$

Since  $M$  is arbitrary, choosing  $M = \frac{2}{\|\beta\|}$  converts (19) into a convex optimization problem

$$\begin{aligned} &\text{minimize}_{\beta, \beta_0} \quad \frac{1}{2} \|\beta\| \\ &\text{subject to} \quad y_i (\beta^\top x_i + \beta_0) \geq 1, \quad \forall x_i \in X, y_i \in Y. \end{aligned} \quad (20)$$

In order to use the primal-dual gradient method proposed in Section III, we need the cost function to be twice differentiable. But, the cost function  $\frac{1}{2} \|\beta\| \notin C^2$ . The optimal solution  $(\beta^*, \beta_0^*)$  of (20), is further equivalent to the optimal solution of

$$\begin{aligned} &\text{minimize}_{\beta, \beta_0} \quad \frac{1}{2} \|\beta\|^2 \\ &\text{subject to} \quad y_i (\beta^\top x_i + \beta_0) \geq 1, \quad \forall x_i \in X, y_i \in Y. \end{aligned} \quad (21)$$

We now use this convex optimization formulation for support vector machines, and derive its primal-dual gradient dynamics.

*Continuous time primal-dual gradient dynamics:* Comparing with the convex optimization formulation given in (1), the cost function is  $f(\beta) = \frac{1}{2} \|\beta\|^2$  and inequality constraints are  $g_i(\beta, \beta_0) = 1 - y_i (\beta^\top x_i + \beta_0)$ ,  $i \in \{1, \dots, 2N\}$ . The Lagrangian can be written as

$$L(\beta, \mu) = \frac{1}{2} \|\beta\|^2 + \sum_{i=1}^{2N} g_i(\beta, \beta_0) \mu_i \quad (22)$$

where  $\mu = (\mu_1, \dots, \mu_{2N})$  denotes the Lagrange variable corresponding to the inequality constraints  $g =$

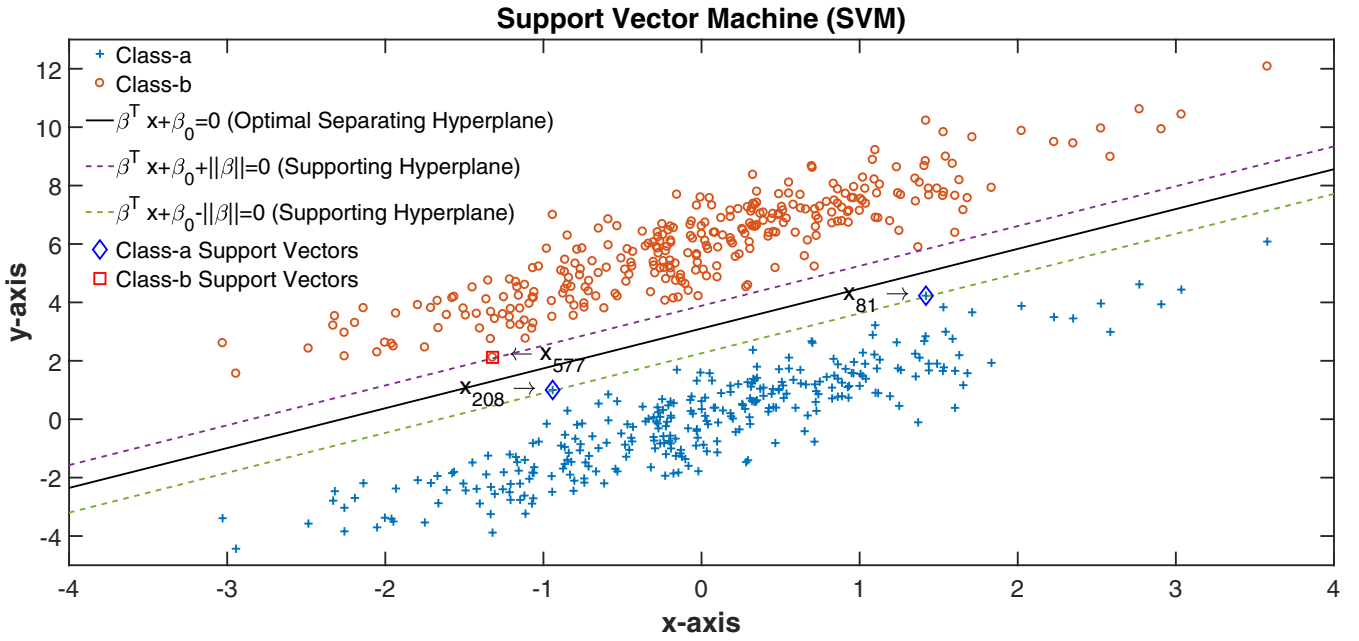


Fig. 5. Classification using Support Vector Machine

$(g_1, \dots, g_{2N})$ . The primal dual gradient laws given in (7) for the convex optimization problem (21) are

$$\begin{aligned} -\tau_\beta \dot{\beta} &= \frac{\partial L}{\partial \beta} \\ -\tau_{\beta_0} \dot{\beta}_0 &= \frac{\partial L}{\partial \beta_0} \\ \tau_{\mu_i} \dot{\mu}_i &= \left( \frac{\partial L}{\partial \mu_i} \right)_{\mu_i}^+ \quad \forall i \in \{1, \dots, 2N\} \end{aligned}$$

equivalently ,

$$\begin{aligned} -\tau_\beta \dot{\beta} &= \beta - \sum_{i=1}^{2N} \mu_i y_i x_i \\ -\tau_{\beta_0} \dot{\beta}_0 &= -\sum_{i=1}^{2N} \mu_i y_i \\ \tau_{\mu_i} \dot{\mu}_i &= (g_i(\beta, \beta_0))_{\mu_i}^+ \quad \forall i \in \{1, \dots, 2N\} \end{aligned} \quad (23)$$

Note that the equilibrium point of the above dynamical system (23) represents the KKT conditions of the optimization problem (21). The first two equations represent the KKT conditions with respect to primal variables and the third equation represents the complementary conditions for the dual variables. This implies, finding the solution of the equilibrium point is equivalent to solving the KKT conditions, which is not a trivial task in many cases. Hence, the equilibria of the above dynamical system is not explicitly known but is implicitly characterized by the optimization problem. Instead of solving for these equilibrium points manually, one can run the dynamical system and use its steady state behavior (points). But to quantify it mathematically, we first have to prove that the dynamical system is globally asymptotically stable at that equilibrium point. To do that we leverage the

propositions presented in the previous sections. We now have the following result.

**Proposition 5.** *The primal-dual dynamics (23) converges asymptotically to the optimal solution of (21).*

*Proof.* Since the optimization problem (21) has a strictly convex cost function and convex inequality constraints, the result follows from Propositions 1 - 4.  $\square$

#### A. Simulation Results

A simulation study is conducted by generating two sets of linearly separable classes having 300 points each, using Normal distribution (see Table I for distribution parameters). Figure 6 presents the evolution of  $\beta$ ,  $\beta_0$ . At equilibrium, the

TABLE I  
 DISTRIBUTION PARAMETERS

	mean	Variance	No. of data points
<b>Class-a</b>	[0 0]	$\begin{bmatrix} 1 & 1.5 \\ 1.5 & 3 \end{bmatrix}$	300
<b>Class-b</b>	[0 6]	$\begin{bmatrix} 1 & 1.5 \\ 1.5 & 3 \end{bmatrix}$	300

primal-dual dynamics in equation (23) results in

$$\beta^* = \sum_{i=1}^{2N} \mu_i^* y_i x_i.$$

**Remark 5.** The results depicted in Figure 8 show that the value of the Lagrange variables, except  $(\mu_{81}, \mu_{208}, \mu_{577})$  are identically equal to zero at equilibrium. Moreover, the data points  $x_{81}$ ,  $x_{208}$  and  $x_{577}$  corresponding to these non-zero Lagrange variables are called support vectors, as can be seen in Fig. 5. The lines passing through these points and parallel to the separating hyperplane are called supporting hyperplanes.

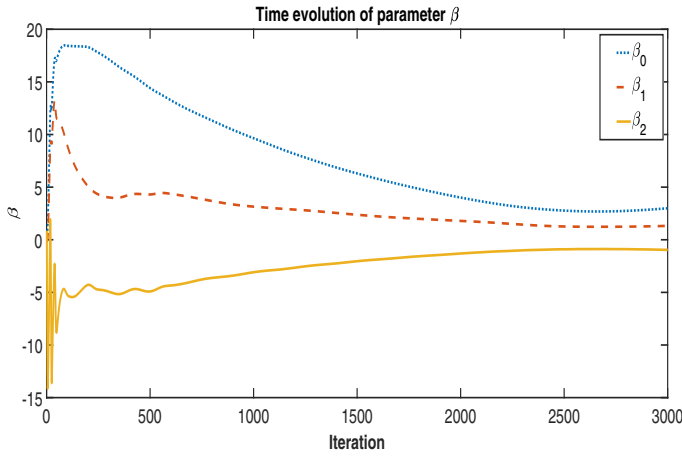


Fig. 6. Time evolution of  $\beta$  and  $\beta_0$

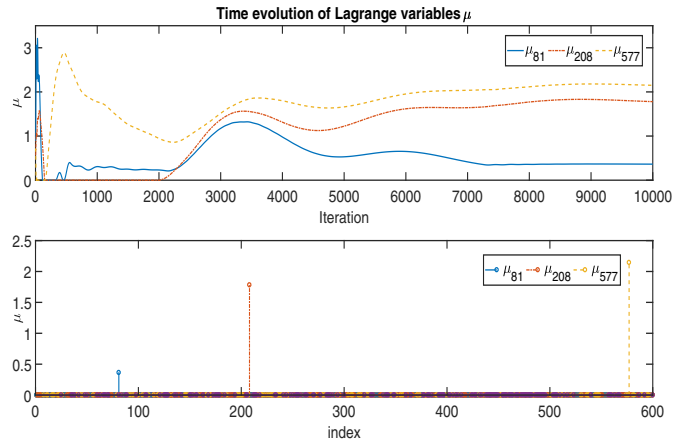


Fig. 8. Time evolution of Lagrange variables  $\mu_i, i \in \{1 \dots 600\}$ .

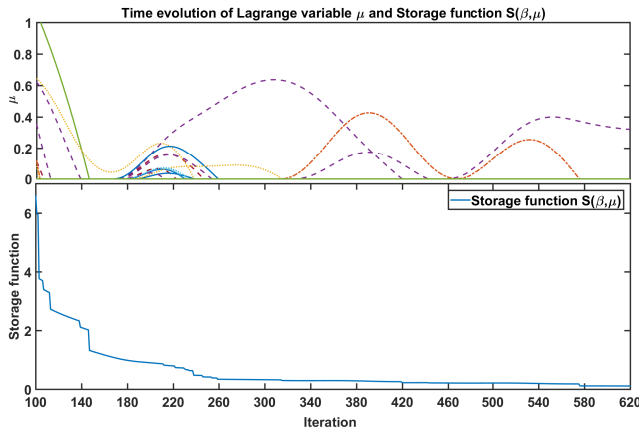


Fig. 7. Time evolution of closed-loop storage function.

Hence

$$\beta^* = \mu_{81}^* x_{81} + \mu_{208}^* x_{208} - \mu_{577}^* x_{577}$$

where the data points  $(x_{81}, x_{208}, x_{577})$  corresponding to these non zero Lagrange variables are support vectors. This implies that the support vectors completely determines the optimal separating hyperplane  $\beta^T x + \beta_0 = 0$  that separates class- $a$  and class- $b$  (see Fig. 7). However, note that one needs to solve the optimization problem, to find these support vectors.

**Remark 6.** Figure 7 shows that, whenever, an inequality constraint becomes feasible (i.e.  $g_i(\beta, \beta_0) \leq 0$ ) and its corresponding Lagrange variable  $\mu_i$  converges to zero, then the closed loop storage function (14) switches to a new storage function that is strictly less than the current one, causing a discontinuity. This is coherent with the Proposition 2, where passivity property is defined with ‘multiple storage functions’.

## V. FUTURE WORK

In Section III, the primal-dual algorithm is treated as interconnected passive systems, (i) convex optimization problem

with only equality constraint, (ii) a state dependent switching system for inequality constraint. Recall that in Proposition 4, we interconnected these systems using

$$\begin{bmatrix} u \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} y \\ \tilde{y} \end{bmatrix} + \begin{bmatrix} v \\ \tilde{v} \end{bmatrix} \quad (24)$$

where  $v$  and  $\tilde{v}$  are considered as new input port-variables of the interconnected system. We can use these new port-variables to analyze and improve the primal-dual gradient laws. The following are some of the important ideas that can be leveraged for future work.

**Robustness:** To analyze uncertainties in parameters or disturbances such as the numerical error accumulated in the primal and dual variables, one can rewrite interconnection as

$$u = \tilde{y} + \Delta \tilde{y} \quad \text{and} \quad \tilde{u} = x + \Delta x \quad (25)$$

where  $\Delta x$  and  $\Delta \tilde{y}$  denotes the numerical error in  $x$  (primal variable) and  $\tilde{y}$  (a function of dual variable) respectively. These can be treated as external disturbances creeping in through the interconnected port variables. We can provide robustness analysis quantitatively (on sensitivity of the algorithm due to numerical errors), using input/output dissipative properties [13] of these systems.

**Stochastic gradient descent:** In SVM simulation we have seen that there are 600 inequality constraints (each corresponds to a data-point). Usually, real world examples may contain many more data-points. Each data-points gives rise to an inequality constraint, and further leads to a gradient-law. In situations involving large data, it is computationally ineffective to run gradient-descent algorithm using all the data-points. In general this obstacle is circumvented using a variation in gradient descent method called *stochastic gradient descent*. Can we propose a passivity based convergence analysis for stochastic gradient descent?

**Control synthesis:** Using these new port variables one can interconnect the primal-dual dynamics to a plant, such that the closed-loop system is again a passive dynamical system [8]. One can also explore the idea of Barrier functions [2] to derive a bounded controller. Gradient methods are inherently distributed computing methods. Hence the controllers

derived from these may inherit this property. Moreover, this framework enables us to solve control problems that whose operating points are characterized by an optimization problems.

#### REFERENCES

- [1] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.
- [2] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [3] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.
- [4] M. Kumar, R. Ghani, and Z.-S. Mei, "Data mining to predict and prevent errors in health insurance claims processing," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 65–74.
- [5] H. Nguyen, M. Butler, A. Roychoudhry, A. Shannon, J. Flack, and P. Mitchell, "Classification of diabetic retinopathy using neural networks," in *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, vol. 4. IEEE, 1996, pp. 1548–1549.
- [6] H. K. Khalil, "Nonlinear systems," *Prentice-Hall, New Jersey*, vol. 2, no. 5, 1996.
- [7] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems," in *Proceedings of the 27th IEEE Conference on Decision and Control*, Dec 1988, pp. 799–803 vol.1.
- [8] T. Stegink, C. De Persis, and A. van der Schaft, "A unifying energy-based approach to stability of power grids with market dynamics," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2612–2622, 2017.
- [9] D. Feijer and F. Paganini, "Stability of primal–dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
- [10] K. C. Kosaraju, V. Chinde, R. Pasumarthy, A. Kelkar, and N. M. Singh, "Stability analysis of constrained optimization dynamics via passivity techniques," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 91–96, Jan 2018.
- [11] T. Kose, "Solutions of saddle value problems by differential equations," *Econometrica, Journal of the Econometric Society*, pp. 59–70, 1956.
- [12] W. Lohmiller and J.-J. E. Slotine, "On contraction analysis for nonlinear systems," *Automatica*, vol. 34, no. 6, pp. 683–696, 1998.
- [13] A. J. van der Schaft, *L2-gain and Passivity Techniques in Nonlinear Control*. Springer, 2017.
- [14] J. Lygeros, K. H. Johansson, S. N. Simic, J. Zhang, and S. S. Sastry, "Dynamical properties of hybrid automata," *IEEE Transactions on automatic control*, vol. 48, no. 1, pp. 2–17, 2003.
- [15] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.
- [16] K. Kosaraju, R. Pasumarthy, N. Singh, and A. Fradkov, "Control using new passivity property with differentiation at both ports," *Indian Control Conference (ICC)*, pp. 7–11, 2017.
- [17] M. Zefran, F. Bullo, and M. Stein, "A notion of passivity for hybrid systems," *IEEE Conference on Decision and Control (CDC)*, 2001.
- [18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.