2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

A Smooth Vector Field for Saddle Point Problems

Hans-Bernd Dürr and Christian Ebenbauer

Abstract— **In this paper we propose a novel smooth vector field whose trajectories globally converge to the saddle point of the Lagrangian associated with a convex and constrained optimization problem. Under suitable assumptions, we prove global convergence of the trajectories for the class of strictly convex problems and we propose a vector field for linear programs.**

I. INTRODUCTION

In this paper we propose a novel way to compute saddle points that arise in convex optimization problems of the form

$$\begin{aligned}&\inf_x f(x)\\&\text{s.t. } g_i(x) \leq 0, i = 1, \ldots, m\end{aligned} \tag{1}$$

with $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, f, g_i convex. Under the assumption of strict feasibility (Slater's Condition), i.e. there exists an $\tilde{x}$ with $g_i(\tilde{x}) < 0$ for $i = 1, \ldots, m$, x^* is the solution of (1) if and only if (x^*, λ^*) is a saddle point of the Lagrangian, i.e. $L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$ for all $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m_+$ with

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x). \tag{2}$$

We propose a continuous-time dynamical system to find a saddle point of the Lagrangian (2). The idea of using a dynamical system to find a saddle point of a Lagrangian function goes back to K. J. Arrow, L. Hurwicz and H. Uzawa as well as to G. W. Brown and J. von Neumann who proposed in [2] (see also [17]) and [7] respectively, a gradient-like system. In order to assure the satisfaction of the constraints and the positivity of the Lagrange multipliers, a method involving a projected gradient in the vector field, i.e.

$$\dot{x} = -\left(\frac{\partial L(x, \lambda)}{\partial x}\right)^\top = -\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x) \tag{3a}$$

$$\dot{\lambda}_i = \mathcal{P}(\lambda_i, g_i(x)), i = 1, \ldots, m, \tag{3b}$$

where the operator $\mathcal{P}(\lambda_i, g_i(x)) := 0$ if $\lambda_i = 0$ and $g_i(x) < 0$, and $\mathcal{P}(\lambda_i, g_i(x)) := g_i(x)$ otherwise, was used to project the vector field to the cone of positive vectors in $\mathbb{R}^m_+$. The projection as it was done in the Arrow-Hurwicz-Uzawa flow (AHU-flow) in (3) renders the vector field non-smooth. Since the work of Arrow, Hurwicz and Uzawa, a lot of further research in the direction of saddle point algorithms has been done, especially in the areas of economics, optimization and game theory. For example, the so-called best response dynamics [4], [14] is a continuous-time dynamical system of the form $\dot{x} \in BR_1(y) - x, \dot{y} \in BR_2(x) - y$ which converges to a saddle point of a zero sum game. While this system is of interest for several reasons, it does not directly yield to an easy implementable algorithm for finding saddle points, since the evaluation of the best response $BR_1(\cdot)$ is an optimization problem itself.

Increasing interest in distributed control and optimization has recently motivated more research in the direction of saddle point algorithms (see [10] and references therein). For example, in [20] a discrete-time primal-dual subgradient algorithm for saddle point problems has been proposed which builds upon the AHU-flow. As pointed out in [20], primal-dual methods are especially appealing in distributed optimization over networks when the dual function cannot be evaluated efficently. The algorithm proposed in this work is also a primal-dual method, but it is in the following aspects different from the work outlined above.

Firstly, in contrast to the AHU-flow, we propose a new method to assure positivity of the Lagrange multipliers by avoiding the (non-smooth) projection operator. Moreover, our method can easily be extended to other cases, e.g. when the Lagrange multipliers are elements in the positive semidefinite matrix cone. Furthermore, the overall idea behind the proposed dynamics is different from the AHU-flow and implicitly involves the dual function. Concerning convergence, we establish global convergence to a saddle point under comparable assumptions as in the AHU-flow. Additionally, the proposed dynamics shows promising results for linear programs and guarantees global asymptotic stability in that case.

Secondly, in contrast to the best response dynamics, the evaluation of the right hand-side involves no optimization problem and is thus easy to implement. What is needed, however, is the evaluation of a Hessian, which is dropped in the linear programming case.

Finally, in contrast to the work of Nedić and Ozdaglar (see [20]), our algorithm is formulated in continuous-time. While discrete-time formulations are as powerful as a continuous-time formulation and often more convenient to implement on a computer, it is on the other hand often advantageous to work in continuous-time, since many powerful mathematical concepts and tools can be applied. Moreover, a differential equation is independent of its algorithmic realization, whereas an algorithm given in the form $x_{k+1} = x_k + \alpha f(x_k)$ has a fixed realization.

Summarizing, our main contribution in this paper is a dynamical system, given as differential equation, which

This project was supported by Deutsche Forschungsgemeinschaft (Emmy-Noether-Grant, Novel Ways in Control and Computation, EB 425/2-1).

Hans-Bernd Dürr and Christian Ebenbauer are with the Institute for Systems Theory and Automatic Control, University of Stuttgart, Germany {hans-bernd.duerr,ce}@ist.uni-stuttgart.de

complements the more dominating literature on discrete-time saddle point algorithms and provides a novel, alternative approach to the AHU-flow with guaranteed convergence properties by avoiding the use of projection operators in the vector field. We prove *global* convergence of the trajectories to a - not necessarily unique - saddle point for a certain class of constrained optimization problems and propose a modified version for linear programs.

The remainder of this paper is structured as follows. In Section II we introduce the notation and the necessary mathematical preliminaries. In Section III we state our main result and the corresponding proofs. In Section IV we demonstrate some aspects of the algorithm by examples. Finally, in Section V we conclude with a summary and outlook.

II. Background

We will make use of the following notation:

The norm of a vector $v \in \mathbb{R}^n$ will be denoted by $||v|| = \sqrt[2]{\sum_{i=1}^n v_i^2}$. A matrix A is said to be positive semidefinite (definite), or $A \geq 0$ ($A > 0$), if it is symmetric $A = A^\top$, and all its eigenvalues are nonnegative (positive). A matrix H is Hurwitz if the real parts of its eigenvalues are in the complex open left half-plane.

$\mathbb{R}^m_+$ denotes the cone of nonnegative vectors in $\mathbb{R}^m$ and $\mathbb{R}^m_{++}$ the interior of $\mathbb{R}^m_+$, i.e. the set of positive vectors in $\mathbb{R}^m$. Define the operator diag : $\mathbb{R}^n \to \mathbb{R}^{n\times n}$ that takes a vector and maps it to a matrix with the elements of the vector on its main diagonal and define the open ball $B_\epsilon(x_0)$ as the set $B_\epsilon(x_0) = \{x \in \mathbb{R}^n : ||x - x_0|| < \epsilon\}$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a C^1- and a C^2-function, respectively. Then

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x}\right)^\top = \left[\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n}\right]^\top$$

$$\nabla L(x,\lambda) = \left(\frac{\partial L(x,\lambda)}{\partial x}\right)^\top = \left[\frac{\partial L(x,\lambda)}{\partial x_1}, \dots, \frac{\partial L(x,\lambda)}{\partial x_n}\right]^\top$$

$$\nabla^2 L(x,\lambda) = \frac{\partial^2 L(x,\lambda)}{\partial x^2}.$$

We continue with some standard results from convex optimization. The following three statements are equivalent: f convex $\Leftrightarrow \nabla f(x)^\top (y-x) \leq f(y) - f(x) \Leftrightarrow \nabla^2 f(x) \geq 0$. Furthermore we have $\forall x \neq y$: f strictly convex $\Leftrightarrow$ $\nabla f(x)^\top (y - x) < f(y) - f(x)$ and on the other hand, a positive definite Hessian $\nabla^2 f(x) > 0$ implies strict convexity of f.

A saddle point of (2) delivers always a solution to (1). In order to guarantee that the converse is true, it suffices to guarantee strict feasibility (Slater's Condition).

The following theorems can be found in the standard literature dealing with optimization. See for example [5], [9] and [21].

Theorem 1: Suppose Slater's Condition is satisfied and f, g_i for $i = 1,\dots,m$ are convex. Then x^* is a solution of (1) if and only if there exists a λ^* such that (x^*, λ^*) is a saddle point of (I) in $\mathbb{R}^n \times \mathbb{R}^m_+$, i.e. $(x^*,\lambda^*) = \arg\inf_{x\in\mathbb{R}^n} \sup_{\lambda\in\mathbb{R}^m_+} L(x,\lambda)$.

Another closely related result is delivered by the Kuhn-Karush-Tucker Theorem.

Theorem 2: Suppose Slater's Condition is satisfied and $f, g_i \in C^1$ for $i = 1,\dots,m$ are convex. Then x^* is a solution of (1) if and only if there exists a $\lambda^* \in \mathbb{R}^m_+$ such that the following three conditions are satisfied:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0, \tag{4a}$$

$$\lambda_i^* g_i(x^*) = 0, i = 1,\dots,m, \tag{4b}$$

$$g_i(x^*) \leq 0, i = 1,\dots,m. \tag{4c}$$

The conditions are also often referred to as *KKT-Conditions*.

III. Main Results

We propose the following equations to find a saddle point of the Lagrangian (2):

$$\dot{x} = -\nabla L(x,\lambda) - \nabla^2 L(x,\lambda)^{-1} \sum_{i=1}^m \lambda_i g_i(x) \nabla g_i(x) \tag{5a}$$

$$\dot{\lambda}_i = \lambda_i g_i(x),\ i = 1,\dots,m \tag{5b}$$

with initial conditions $x(0) = x_0 \in \mathbb{R}^n$ and $\lambda_i(0) = \lambda_{i0} \in \mathbb{R}_{++}$ for $i = 1,\dots,m$.

Before we prove the main results, we briefly discuss the structure of these equations and the underlying ideas.

Firstly, we observe that the equilibrium points of (5a) and (5b) in $\mathbb{R}^n \times \mathbb{R}^m_+$ are the points which satisfy the KKT-Conditions (4a) and (4b).

Secondly, comparing the equations with the AHU-flow (3), we see a modification in the λ-dynamics in (5b). The idea behind (5b) is to guarantee non-negativity of the Lagrange multipliers in a natural way and to avoid switching (projection) as it was done in the AHU-flow. Notice that the λ_i's stay non-negative if initialized in $\mathbb{R}^m_{++}$. It is also not difficult to see that (5b) can be generalized to other conic constraints, e.g. when dealing with positive semidefinite Lagrange multiplier matrices. In that case one can imagine to replace (5b) by the matrix equation $\dot{\Lambda} = \frac{1}{2}(G^\top \Lambda + \Lambda G)$ with $\Lambda > 0$ (for further information on dynamical systems involving matrices see e.g. [1] and [13]).

Thirdly, we see an additional expression on the right hand side in the x-dynamics (5a). This expression arises from the following idea which leads to the system in (5). The task of the x-dynamics (5a) is to minimize $L(x,\lambda)$. If $\lambda(t)$ was constant, then the right hand side of (5a) would just consist of the negative gradient of $L(x,\lambda)$, but since $\lambda(t)$ itself is evolving, its dynamics have to be taken into account, which leads to the additional term $-\nabla^2 L(x,\lambda)^{-1} \sum_{i=1}^m \lambda_i g_i(x) \nabla g_i(x)$.

The overall idea underlying the dynamics of system (5) is illustrated in Fig. 1. The system is designed in such a way that the x-dynamics will minimize $L(x,\lambda)$ with respect to x, i.e. by zeroing

$$V_1(x,\lambda) = \frac{1}{2} \nabla L(x,\lambda)^\top \nabla L(x,\lambda) \tag{6}$$

while the λ-dynamics will maximize the (dual) function

$$V_2(\lambda) = -L(\bar{x}(\lambda),\lambda) \tag{7}$$

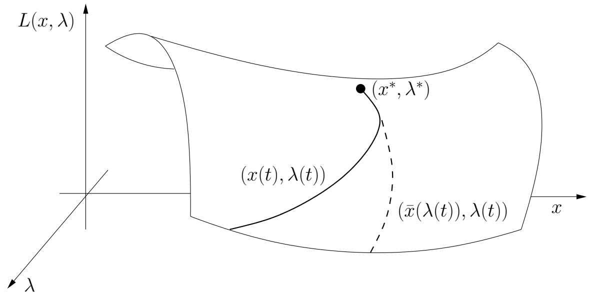

Fig. 1: Tracking of $\bar{x}(\lambda)$

on $\mathbb{R}_+^m$, with $\bar{x}(\lambda)$ as the minimizer x of $L(x,\lambda)$ with respect to some $\lambda \in \mathbb{R}_+^m$. In the following, we prove convergence to a saddle point of $L(x,\lambda)$ in (2), assuming that Slater's Condition is satisfied, f, g_i's are strictly convex and possess a minimum.

Hereby, we proceed in several steps. In Lemma 1 we show that $\bar{x}_c(\lambda)$ (see (8), (9), (10)) with respect to $\lambda \in \mathbb{R}_+^m$ is always bounded. In Lemma 2 we establish radially unboundedness of $V_1(x,\lambda)$ with respect to x by introducing an auxiliary optimization problem. Lemma 3 states that under the assumption of forward completeness of $(x(t),\lambda(t))$, $x(t)$ converges to the set $V_1(x,\lambda)=0$ whereas the existence and boundedness of $x(t)$ and $\lambda(t)$ is established in Lemma 4. Finally in Theorem 3 we prove convergence of trajectories to a saddle point.

Global convergence of the trajectories to a saddle-point is established under the assumptions (A1) – (A3).

Assumptions:

(A1) f, g_i, $i=1,\dots,m$, are twice continuously differentiable and strictly convex. Moreover $\nabla^2 f(x) > 0$.

(A2) The functions $f_c(x) = f(x) - c^\top x, g_i$, $i=1,\dots,m$ possess a minimum for any $c \in \mathbb{R}^n$.

(A3) $\exists \tilde{x} \in \mathbb{R}^n$ such that $g_i(\tilde{x}) < 0$, $i=1,\dots,m$.

Assumption (A2) is necessary since strict convexity of a function does not imply compact level sets, as e.g. the function $-\ln(x)$ is strictly convex, but does not admit a minimum. Since we need compact level sets, we restrict f and all g_i's to be strictly convex and additionally f_c and all g_i's must admit a minimum. The additional linear term $-c^\top x$ appearing in f_c is a technical assumption needed in the proof and assures, loosely speaking, that f grows faster than a linear function.

We now introduce an auxiliary optimization problem, whose link to problem (1) will be clear immediately.
Consider the following optimization problem

$$\begin{aligned} &\inf_x \; f(x) - c^\top x \\ &\text{s.t. } g_i(x) \le 0, i=1,\dots,m \end{aligned} \tag{8}$$

with $c \in \mathbb{R}^n$. Observe that the Lagrangian

$$L_c(x,\lambda) = f(x) - c^\top x + \sum_{i=1}^m \lambda_i g_i(x) \tag{9}$$

associated with problem (8) is strictly convex in x. Before we state our main results, we introduce the quantity $\bar{x}_c(\lambda)$ that is implicitly defined by

$$\nabla L_c(\bar{x}_c(\lambda),\lambda) = \nabla f(\bar{x}_c(\lambda)) - c + \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x}_c(\lambda)) = 0. \tag{10}$$

Obviously, the Lagrangian $L(x,\lambda)$ in (2) associated with problem (1) coincides with the Lagrangian $L_c(x,\lambda)$ in (9) for $c=0$, i.e. $L(x,\lambda) = L_0(x,\lambda)$. Therefore, we can also state that $\bar{x}(\lambda)$ as the minimum of $L(x,\lambda)$ for some $\lambda \in \mathbb{R}_+^m$ coincides with $\bar{x}_0(\lambda)$, i.e. $\bar{x}(\lambda) = \bar{x}_0(\lambda)$.

$L_c(x,\lambda)$ is strictly convex in x. This implies that $\bar{x}_c(\lambda)$ is the unique solution of the implicit equation (10) for any $\lambda \in \mathbb{R}_+^m$ and $c \in \mathbb{R}^n$. Due to assumption (A2) and the *Implicit Function Theorem*, $\bar{x}_c(\lambda)$ is finite for any finite $\lambda \in \mathbb{R}_+^m$ and is a C^1-function. Its derivative is given by (see Appendix VI-A) $\frac{\partial \bar{x}_c(\lambda)}{\partial \lambda} = -\left(\nabla^2 L(\bar{x}_c(\lambda),\lambda)\right)^{-1}\left[\nabla g_1(\bar{x}_c(\lambda)),\dots,\nabla g_m(\bar{x}_c(\lambda))\right]$.

Lemma 1: Let assumptions (A1) – (A3) be satisfied and suppose $\bar{x}_c(\lambda)$ is the unique minimizer of $L_c(x,\lambda)$ for a given $\lambda \in \mathbb{R}_+^m$ and $c \in \mathbb{R}^n$. Then $\sum_{i=1}^m \lambda_i g_i(\bar{x}_c(\lambda)) < 0$ whenever $\sum_{i=1}^m \lambda_i$ is sufficiently large. In particular, $\exists j \in \{1,\dots,m\}$ such that $g_j(\bar{x}_c(\lambda)) < 0$.

Proof: We prove the result by contradiction. Suppose $\sum_{i=1}^m \lambda_i g_i(\bar{x}_c(\lambda)) \ge 0$ whenever $\sum_{i=1}^m \lambda_i$ is sufficiently large. By optimality of $\bar{x}_c(\lambda)$, we must have $L_c(\bar{x}_c(\lambda),\lambda) \le L_c(x,\lambda)$, $\forall x \ne \bar{x}_c(\lambda)$ and by assumption (A3) $\exists \tilde{x}$ such that $g_i(\tilde{x}) < 0, \forall i \in \{1,\dots,m\}$. Then $L_c(\bar{x}_c(\lambda),\lambda) - L_c(\tilde{x},\lambda) = f(\bar{x}_c(\lambda)) - c^\top \bar{x}_c(\lambda) + \sum_{i=1}^m \lambda_i g_i(\bar{x}_c(\lambda)) - f(\tilde{x}) + c^\top \tilde{x} - \sum_{i=1}^m \lambda_i g_i(\tilde{x}) \ge f(\bar{x}_c(\lambda)) - c^\top \bar{x}_c(\lambda) - f(\tilde{x}) + c^\top \tilde{x} - \max_k g_k(\tilde{x}) \sum_{i=1}^m \lambda_i$.

By assumptions (A1) and (A2) $\tilde{x}$ lies in a compact set, furthermore $f(\bar{x}_c(\lambda)) - c^\top \bar{x}_c(\lambda)$ and $f(\tilde{x}) - c^\top \tilde{x}$ are bounded from below, since by (A2) they possess a minimum. We see that $L_c(\bar{x}_c(\lambda),\lambda) - L_c(\tilde{x},\lambda) > 0$ whenever $\sum_{i=1}^m \lambda_i$ has been chosen sufficiently large. This contradicts the assumption that $\sum_{i=1}^m \lambda_i g_i(\bar{x}_c(\lambda)) \ge 0$ whenever $\sum_{i=1}^m \lambda_i$ is large.

Since $\sum_{i=1}^m \lambda_i g_i(\bar{x}_c(\lambda)) < 0$ and $\lambda_i \ge 0$, we conclude that $\exists j \in \{1,\dots,m\}$ such that $g_j(\bar{x}_c(\lambda)) < 0$. ■

In order to use the previously defined Lyapunov function $V_1(x,\lambda)$ in (13), we first show that it is radially unbounded in x, i.e. $V_1(x,\lambda) \to \infty$ whenever $||x|| \to \infty$. For this purpose, we introduce a set containing all x belonging to the level sets of $V_1(x,\lambda)$ and prove that this set is always compact. Let M_α denote the set $M_\alpha = \{x \in \mathbb{R}^n : ||\nabla L(x,\lambda)|| = \alpha, \lambda \in \mathbb{R}_+^m\}$.

Lemma 2: Let assumptions (A1) – (A3) be satisfied, then for any fixed $\alpha \ge 0$ the set M_α is nonempty and compact. Furthermore, $V_1(x,\lambda)$ is radially unbounded in x.

Proof: We will first proof that the set

$$\begin{aligned} N_c &= \{x \in \mathbb{R}^n : x = \bar{x}_c(\lambda), \lambda \in \mathbb{R}_+^m\} \\ &= \{x \in \mathbb{R}^n : \nabla L(x,\lambda) = c, \lambda \in \mathbb{R}_+^m\} \end{aligned} \tag{11}$$

is compact. In order to prove that N_c is compact, we will make use of the fact that $\bar{x}_c(\lambda)$ defined by (10) is unique and exists for all $\lambda \in \mathbb{R}_+^m$. By assumptions (A1) and (A2), $\bar{x}_c = \bar{x}_c(\lambda)$ is a well-defined continuous function and therefore there exists a unique and bounded $\bar{x}_c(\lambda)$ for $\sum_{i=1}^m \lambda_i$ being in a compact set. Hence, it remains to show that $\bar{x}_c = \bar{x}_c(\lambda)$ is bounded whenever $\sum_{i=1}^m \lambda_i \to \infty$. This directly follows from Lemma 1, since for $\sum_{i=1}^m \lambda_i$ sufficiently large $\exists j \in \{1, \dots, m\}$ s.t. $g_j(\bar{x}_c(\lambda)) < 0$ and since the functions g_i admit compact level sets due to assumptions (A1) and (A2), $\bar{x}_c(\lambda)$ must be bounded.

Observe now that $M_\alpha = \bigcup_{||c||=\alpha} N_c$, since

$$
\begin{aligned}
&\tilde{x} \in \{x \in \mathbb{R}^n : ||\nabla L(x,\lambda)|| = \alpha, \lambda \in \mathbb{R}^m_+\} \\
&\text{ if and only if} \\
&\tilde{x} \in \bigcup_{||c||=\alpha} \{x \in \mathbb{R}^n : \nabla L(x,\lambda) = c, \lambda \in \mathbb{R}^m_+\}
\end{aligned} \tag{12}
$$

and since $||c|| = \alpha$ defines a compact set, we conclude that M_α is nonempty and compact.

Finally, we prove that $V_1(x,\lambda)$ admits compact level sets in x. Equivalently to radially unboundedness, i.e. $||x|| \to \infty \Rightarrow V_1(x,\lambda) \to \infty$, we can also write $V_1(x,\lambda) \leq \alpha \Rightarrow ||x|| \leq \gamma$. The latter follows immediately from the obtained result, since M_α denotes the set of x contained in a level set of $V_1(x,\lambda)$ and as M_α is bounded for all $\lambda \in \mathbb{R}^m_+$ and any fixed $\alpha \geq 0$, we conclude that $V_1(x,\lambda)$ is radially unbounded in x. ■

We are now ready to prove the existence and boundedness of $x(t)$ and $\lambda(t)$ using the Lyapunov function $V_1(x,\lambda)$. A crucial issue in the proof is the existence of trajectories, especially of $\lambda(t)$, as $V_1(x,\lambda)$ is only radially unbounded in x, i.e. a negative derivative of $V_1(x,\lambda)$ along the trajectories of (5) does not imply boundedness of $\lambda(t)$ but only of $x(t)$.

Lemma 3: Let assumptions (A1) – (A3) be satisfied. Moreover, suppose a solution of $(x(t),\lambda(t))$ of (5) initialized in $\mathbb{R}^n \times \mathbb{R}^m_{++}$ is forward complete. Then $\lim_{t\to\infty} ||\nabla L(x(t),\lambda(t))|| \to 0$, equivalently $\lim_{t\to\infty} ||x(t) - \bar{x}(\lambda(t))|| = 0$.

Proof: Consider the Lyapunov function $V_1(x,\lambda) = \frac{1}{2}\nabla L(x,\lambda)^\top \nabla L(x,\lambda)$ that is by Lemma 2 radially unbounded in x.

The derivative along the trajectories of (5) yields

$$
\begin{aligned}
\dot{V}_1 &= (\nabla L(x,\lambda))^\top \left(\nabla^2 L(x,\lambda)\dot{x} + \sum_{i=1}^m \dot{\lambda}_i \nabla g_i(x)\right) \\
&= -\nabla L(x,\lambda)^\top \nabla^2 L(x,\lambda) \nabla L(x,\lambda).
\end{aligned} \tag{13}
$$

Since $\nabla^2 L(x,\lambda)$ is positive definite and $x(t)$ is forward complete by assumption, $x(t)$ converges to $\nabla L(x,\lambda) = 0$. We conclude that $x(t)$ approaches $\bar{x}(\lambda(t))$ as $t \to \infty$ and furthermore $\lim_{t\to\infty} ||\nabla L(x,\lambda)|| \to 0$ which implies $\lim_{t\to\infty} x(t) \in M_0 = N_0$. ■

Lemma 4: Let assumptions (A1) – (A3) be satisfied. Then any solution $(x(t),\lambda(t))$ of (5) initialized in $\mathbb{R}^n \times \mathbb{R}^m_{++}$ is forward complete and bounded.

Proof: First, we show that the solution $(x(t),\lambda(t))$ is forward complete. Due to Lemma 3, $x = x(t)$ is bounded over the maximal interval of existence of $\lambda = \lambda(t)$ since $h(t) := ||\nabla L(x(t),\lambda(t))||$ is bounded and monotonically decreasing and thus $x(t) \in M_{h(t)}$ is always bounded as long as the solution $\lambda(t)$ exists. Therefore, the solution is forward complete (bounded), if $\lambda = \lambda(t)$ exists for all $t \geq 0$ (is bounded).

The existence of $\lambda = \lambda(t)$ follows by integrating (5b), that yields $\lambda_i(t) = \lambda_{i0} e^{\int_0^t g_i(x(\tau))d\tau}, i = 1, \dots, m$.

Since $x(\tau)$ stays bounded for $\tau \in [0,t]$, the integral $\int_0^t g_i(x(\tau))d\tau$ has a finite value for any $t \geq 0$. Hence, $\lambda(t)$ is forward complete.

Finally we show that $\sum_{i=1}^m \lambda_i(t)$ is bounded. Consider the Lyapunov function $W(\lambda) = \sum_{i=1}^m \lambda_i$ that is positive definite on $\mathbb{R}^m_+$ and radially unbounded. The derivative along the trajectories yields $\dot{W} = \sum_{i=1}^m \dot{\lambda}_i = \sum_{i=1}^m \lambda_i g_i(x)$. Substituting $x = \xi + \bar{x}(\lambda)$, we obtain $\dot{W} = \sum_{i=1}^m \lambda_i g_i(\xi + \bar{x}(\lambda))$. Due to Lemma 1 we have $\sum_{i=1}^m \lambda_i g_i(\bar{x}(\lambda)) < 0$ whenever $\sum_{i=1}^m \lambda_i$ is sufficiently large. Moreover, due to Lemma 3 we have $\lim_{t\to\infty} ||\xi(t)|| = 0$, and with forward completeness of $(x(t),\lambda(t))$, we conclude that there exists a time $t_1 > 0$ such that $\forall t > t_1$ we have $\dot{W} = \sum_{i=1}^m \lambda_i g_i(\xi + \bar{x}(\lambda)) < 0$ whenever $\sum_{i=1}^m \lambda_i$ is sufficiently large. Thus, since $W(\lambda)$ is radially unbounded on $\mathbb{R}^m_+$, $\lambda = \lambda(t)$ stays bounded. ■

As seen before, a crucial issue in this work is to establish forward completeness and boundedness of the solutions. In the proof, the use of assumptions (A2) and (A3) plays an important role. The fact that $\sum_{i=1}^m \lambda_i g_i(\bar{x}(\lambda))$ is *strictly* smaller than zero assures that there exists a *finite* time such that $\dot{W} < 0$.

We are now ready to state our main result, namely the convergence of $(x(t),\lambda(t))$ to a saddle point (x^*,λ^*) of $L(x,\lambda)$ in (2) and therefore to the solution of (1).

Theorem 3: Let assumptions (A1) – (A3) be satisfied, suppose f and g_i, $i = 1, \dots, m$ are analytic functions and let (x^*,λ^*) be the unique saddle point of L. Then any solution of $(x(t),\lambda(t))$ of (5a) and (5b) initialized in $\mathbb{R}^n \times \mathbb{R}^m_{++}$ converges to a single connected component E_i of the set

$$
E = \left\{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+ : x = \bar{x}(\lambda), \lambda_i g_i(x) = 0\right\}. \tag{14}
$$

Furthermore there exist finitely many connected components E_i, where all of them are unstable except the singleton E_0 containing the saddle point (x^*,λ^*) that is asymptotically stable.

Proof: Step 1: We show that all points $(\bar{x}^*,\bar{\lambda}^*)$ with $\bar{x}^* = \bar{x}(\bar{\lambda}^*)$ and $\bar{\lambda}^* = \arg\max_{\lambda\in\mathbb{R}^m_+} L(\bar{x}(\lambda),\lambda)$ are saddle points. This can be verified using the necessary condition for a function having a maximum at $\bar{\lambda}^* \in \mathbb{R}^m_+$ (see [5] , p. 177, Ex. 1.1):

$$
\begin{aligned}
\left(\nabla L(x,\lambda)\frac{\partial \bar{x}(\lambda)}{\partial \lambda} + \frac{\partial L(x,\lambda)}{\partial \lambda}\right)|_{\bar{x}(\bar{\lambda}^*)} &= g_i(\bar{x}(\bar{\lambda}^*)) \leq 0 \\
\bar{\lambda}_i^* g_i(\bar{x}(\bar{\lambda}^*)) &= 0, i = 1, \dots, m.
\end{aligned} \tag{15}
$$

With this and since $x = \bar{x}(\lambda)$ all KKT-Conditions are fulfilled and all $(\bar{x}(\bar{\lambda}^*),\bar{\lambda}^*)$ are saddle points of L. Since

(x^*, λ^*) is unique, the maximum of $L(\bar{x}(\lambda), \lambda)$ is unique and $(\bar{x}^*, \bar{\lambda}^*) = (x^*, \lambda^*)$.

Step 2: We conclude with assumptions (A1) – (A3) and by Lemma 4 that $x(t)$ and $\lambda(t)$ are bounded. Choose $\mathcal{M} := \mathbb{R}^n \times \mathbb{R}^m$ as the manifold with the standard Euclidian metric and define the set $\mathcal{S} = \left\{[x, \lambda]^\top : x = \bar{x}(\lambda), \lambda \in \mathbb{R}^m_+\right\} \cap \Omega$ with Ω a compact set containing the ω-limit set of $(x(t), \lambda(t))$ initialized at $\mathbb{R}^n \times \mathbb{R}^m_{++}$. Obviously, we have $\mathcal{S} \subset \mathcal{M}$. By Lemma 3 we conclude that the set $\mathcal{S}$ is attractive for solutions of (5).

We introduce the height (dual) function $V_2(\lambda) := -L(\bar{x}(\lambda), \lambda)$ and calculate the derivative $\dot{V}_2$ along the trajectories of (5) on the set $\mathcal{S}$. This yields $\dot{V}_2 = -\left(\nabla L(x, \lambda)\frac{\partial \bar{x}(\lambda)}{\partial \lambda} + \frac{\partial L(x,\lambda)}{\partial \lambda}\right)|_{\bar{x}(\lambda)}\dot{\lambda} = -\sum_{i=1}^m g_i(\bar{x}(\lambda))^2 \lambda_i$.
Define the set $E := \{[x, \lambda]^\top : x = \bar{x}(\lambda), \lambda \in \mathbb{R}^m_+, e(\lambda) = 0\}$ with $e(\lambda) : \lambda \mapsto [\lambda_1 g_1^2(\bar{x}(\lambda)), \ldots, \lambda_m g_m^2(\bar{x}(\lambda))]$. Observe that E is the set where $\dot{V}_2 = 0$, i.e. $\dot{V}_2 < 0$ on $\mathcal{S}\backslash E$ and it contains the set of equilibrium points satisfying the *KKT-Conditions* (4a), (4b).

Step 3: Define the set $E_\lambda := \{\lambda \in \mathbb{R}^m_+ : e(\lambda) = 0\}$ and notice that every connected component of E_λ is contained in a level set of $V_2(\lambda)$ if and only if every connected component of E is contained in $V_2(\lambda)$, since $V_2(\lambda) = L(\bar{x}(\lambda), \lambda)$.

In order to apply Theorem 6 in [3] (see Appendix VI-B), we will now show that every connected component of E_λ is contained in a level set of $V_2(\lambda)$. In order to prove this, we show that $V_2(\lambda)$ is constant in directions of the tangent cone $T^C_\lambda E_\lambda$ at any point $\lambda \in E_\lambda$, i.e. $\frac{d}{d\epsilon}V_2(\lambda + \epsilon w)|_{\epsilon=0} = \frac{\partial V_2(\lambda)}{\partial \lambda}w = 0, \forall w \in T^C_\lambda E_\lambda$ with $\frac{\partial V_2(\lambda)}{\partial \lambda} = \frac{\partial L(x,\lambda)}{\partial \lambda}|_{\bar{x}(\lambda)} = [g_1(\bar{x}(\lambda)), \ldots, g_m(\bar{x}(\lambda))]$, because $\nabla L(x, \lambda)|_{\bar{x}(\lambda)} = 0$. Since (see e.g. [12, p. 44]) $T^C_\lambda E_\lambda \subseteq R_\lambda E_\lambda = \{v \in \mathbb{R}^m : \frac{\partial e}{\partial \lambda}|_{\lambda \in E_\lambda} v = 0\}$, it suffices to prove that $\frac{\partial V_2(\lambda)}{\partial \lambda}v = 0, \forall v \in R_\lambda E_\lambda$. This implies that we stay on a level set of $V_2(\lambda)$ for a small perturbation in the direction of $v \in R_\lambda E_\lambda$, and therefore also for any $w \in T^C_\lambda E_\lambda$. The set E_λ contains all points with $\lambda_i g_i^2(\bar{x}(\lambda)) = 0$. We distinguish between two cases:

1) $\forall i \in \{1, \ldots, m\} : g_i(\bar{x}(\lambda)) = 0$.
   It follows that $\frac{\partial V_2(\lambda)}{\partial \lambda} = 0$. Since the dual function is concave ($V_2(\lambda)$ is convex, see [5]) each point where the gradient vanishes is a global maximizer of the dual function, and consequently the connected components where the gradient vanishes, must all lie on the same level set[1].
2) $\exists i \in \{1, \ldots, m\} : g_i(\bar{x}(\lambda)) \neq 0$.
   Without loss of generality, we relabel all $g_i(\bar{x}(\lambda))$ such that:
   - $g_i(\bar{x}(\lambda)) \neq 0$, $i = 1, \ldots, k$
   - $g_i(\bar{x}(\lambda)) = 0$, $i = k + 1, \ldots, m$

   with $1 \leq k \leq m$. Then, the Jacobian of $e(\lambda)$ at $\lambda \in E$ yields:
   $$\frac{\partial e}{\partial \lambda}|_{\lambda \in E} = \begin{bmatrix} \mathrm{diag}[g_1^2(\bar{x}(\lambda)), \ldots, g_k^2(\bar{x}(\lambda))] & 0 \\ 0 & 0 \end{bmatrix}. \quad (16)$$
   Therefore, all vectors in $R_\lambda E_\lambda$ must be of the form:
   $$v = \left[0, \ldots, 0, v_{k+1}, \ldots, v_m\right]^\top, \quad (17)$$
   and the directional derivative of $V_2(\lambda)$ along the elements in $R_\lambda E_\lambda$ yields $\frac{\partial V_2(\lambda)}{\partial \lambda}v = \sum_{i=k+1}^m g_i(\bar{x}(\lambda))v_i = 0, \forall v \in R_\lambda E_\lambda$.

Therefore, all vectors $v \in R_\lambda E_\lambda$ are perpendicular to the gradient on the level set of $V_2(\lambda)$ and since $T^C_\lambda E \subseteq R_\lambda E_\lambda$ all points of a connected component of E lie in a level set of $V_2(\lambda)$. Thus, E_λ is contained in a level set of $V_2(\lambda)$.

The set E can be decomposed such that $E = \bigcup_{i \in I} E_i$ with $\{E_i\}_{i \in I}$ denoting the connected components in E and let $\nu^i \in \mathcal{V} = V_2(E)$ denote the image of of E under $V_2(\lambda)$ (we write $V_2(E)$ instead of $-L(E)$), i.e. $\nu^i = V_2(E_i)$. Since all $\{E_i\}_{i \in I}$ are contained in the level sets of $V_2(\lambda)$, we conclude that *different* ν^i's must posses *different* preimages with $z^i = [x^i, \lambda^i]^\top \in E_i$. Remember that x^i is uniquely defined for a fixed λ^i, since $z^i \in E_i$ with $x = \bar{x}(\lambda)$, therefore $E = \{[x, \lambda]^\top : \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) = 0, \lambda_i g_i(x) = 0, i = 1, \ldots, m\}$, and thus z^i is a *zero of an analytic function*. We will now exclude the existence of accumulation points in $\mathcal{V}$. Suppose, there exists an accumulation point ν^∞, i.e. there exists a sequence $\{\nu^k\}$, with *distinct* ν^k such that (see [19]) $\lim_{k \to \infty} \nu^k = \nu^\infty$. Let z^k denote some preimage of ν^k, i.e. $z^k \in E_k$. The set E is compact, so there exists a converging subsequence $\{\tilde{z}^j\} \subseteq \{z^k\}$ (Thm. of Bolzano-Weierstrass, [19]) with *distinct* $\tilde{z}^j$ such that $\lim_{j \to \infty} \tilde{z}^j = \tilde{z}^\infty$, $\tilde{z}^\infty \in V_2^{-1}(\nu^\infty)$. Since $\tilde{z}^\infty$ is the accumulation point of $\{\tilde{z}^j\}$, there exists for every $\epsilon > 0$ a $p \in \mathbb{N}$ such that $\tilde{z}^j \in B_\epsilon(\tilde{z}^\infty), \forall j > p$, i.e. for any $\epsilon > 0$ there exist *infinitely* many distinct $\tilde{z}^j \in B_\epsilon(\tilde{z}^\infty)$ and therefore *infinitely* many distinct connected components E_j ($\tilde{z}^j \in E_j$) intersect with an ϵ-ball $B_\epsilon(\tilde{z}^\infty)$. This leads to a contradiction since f and g_i, $i = 1, \ldots, m$ are analytic and not identically zero (there exists an x and d such hat $h(\alpha) = f(x + \alpha d) \not\equiv 0$) it follows by Lojasiewicz's Theorem (see Thm. 6.3.3 in [18]) that the zero set of a real valued analytic functions is locally a union of a *finite* number of connected components. This contradicts the assumption that an accumulation point $\tilde{z}^\infty$ exists. We conclude that the connected components in E are isolated and are embedded in the level sets of $V_2(\lambda)$.

We are now ready to apply Theorem 6 in [3] (see Appendix VI-B), concluding that the trajectories of (5) approach a connected component E_i of E with $t \to \infty$.

Observe furthermore that E contains all equilibrium points of (5) which satisfy (4a) and (4b) but not necessarily (4c).

Step 4: We will now show by using Chetaev's Instability Theorem that all E_i are unstable except the singleton E_0 containing the saddle point (x^*, λ^*). Take again $V_2(\lambda) = -L(\bar{x}(\lambda), \lambda)$ and define $W_i(\lambda) = -V_2(\lambda) + c_i$ with $(x^i, \lambda^i) \in E_i, i \geq 1$ and the constant $c_i = L(x^i, \lambda^i)$. Obviously $W_i(\lambda^i) = 0$. Define $\bar{B}_\epsilon := B_\epsilon(x^i, \lambda^i) \cap N_0 \times \mathbb{R}^m_+$,

[1] The convexity condition is needed, since there exists (non-convex and non-concave) functions not constant on a connected set of points where its gradient vanishes (see [22]).

$(x^i, \lambda^i) \in E_i, i \geq 1$ with N_0 as in (11). Since $W_i(\lambda^*) > W_i(\lambda^i)$, $i \geq 1$ and $W_i, i \geq 0$ are concave, it follows that for every ϵ there exists a subset D of $\bar{B}_\epsilon$ such that $W_i(\lambda) > 0$ for all $\lambda \in D \backslash \lambda^i$. Now since N_0 is invariant and $\dot{W}_i = \sum_{i=1}^n g_i(\bar{x}(\lambda))^2 \lambda_i > 0$, for all $(\bar{x}(\lambda), \lambda) \in \bar{B}_\epsilon$ (and therefore also on D) it follows by Chetaev's Instability Theorem (see [16]) that the sets $E_i, i \geq 1$ are unstable.

In the following we use Theorem 11 in [15] in order to show that E_0, i.e. the saddle point, is asymptotically stable. Note that V_1 in (13) is a positive semi-definite Lyapunov function on $\mathbb{R}^n \times \mathbb{R}^m_+$ that vanishes on the set N_0. In order to show that the saddle-point (x^*, λ^*) is asymptotically stable it is sufficient to show that the restriction of (5) is asymptotically stable on the set N_0[2]. Using the Lyapunov function $-W_0$ that is positive definite on N_0 and vanishes only at the saddle-point, the Lie-derivative yields $\dot{W}_0 = -\sum_{i=1}^n g_i(\bar{x}(\lambda))^2 \lambda_i$ and is strictly negative definite for all $\lambda \neq \lambda^*$. This proves that the saddle point $(x^*, \lambda^*) \in E_0$ is asymptotically stable on N_0. Using the results of Theorem 11 in [15] we can conclude from the asymptotic stability of (x^*, λ^*) on N_0 that (x^*, λ^*) is asymptotically stable. ■

In the following, we discuss the main result and the assumption under which it has been established as well as the applicability of (5) to linear programming.

First note that the existence of a unique saddle point of L is assumed to be able to use Theorem 11 in [15], since it is formulated such that it is valid for a point. It can be expected however that this is also valid in the case that E_0 is not a singleton.

Second it is important to point out that the system converges to E_0 for practically every initial condition. The convergence to an unstable connected component E_i is highly unlikely, but possible in some pathological cases.

We restrict the functions f and g_i, $i = 1, \dots, m$ to be analytic because a finite number of accumulation points in $V_2(E)$ is required for Theorem 6 in [3]. Since analytic functions can have *only* isolated connected components, we assure the non-existence of accumulation points. On the other hand, infinitely many accumulation points can only occur in highly degenerated cases. As pointed out in [11], most nonlinear differentiable programs are well-behaved. Furthermore, a *finite* number of accumulation points still leads to a guaranteed convergence but there is no general criteria known to the authors that gives a reasonable characterization of such functions.

Finally, the assumptions (A1) – (A3) are conservative but needed in our current proof, mainly in order to establish boundedness of solutions. From a practical point of view, one might expect that (5) will converge in more general situations when the assumptions are not satisfied.

The application of (5) can be extended to linear programs. The proof in this special case can be established using a single Lyapunov-function and under less strict assumptions as those needed in Theorem 3.

[2] In terms of Theorem 11 in [15]: $V_1 = V$ and $N_0 = L$ with L being the largest invariant set contained in $\{x : \dot{V}(x) = 0\}$.

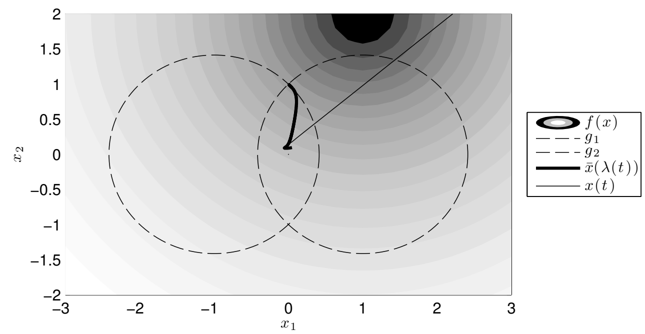

Fig. 2: Feasible Sets and Trajectory $(x_1(t), x_2(t))$

In the following we will present the structure of the dynamical system for linear programs.

First recall a linear program in inequality form:

$$\begin{aligned} &\inf_x \; c^\top x \\ &\text{s.t. } a_i^\top x - b_i \leq 0, i = 1, \dots, m. \end{aligned} \tag{18}$$

We propose the following vector field for this type of problems:

$$\dot{x} = -c - \sum_{i=1}^m \lambda_i a_i (a_i^\top x - b_i + 1) \tag{19a}$$

$$\dot{\lambda}_i = \lambda_i (a_i^\top x - b_i), \; i = 1, \dots, m. \tag{19b}$$

This system is a special realization of (5) for linear programs where $\nabla^2 L(x, \lambda)^{-1}$ was substituted by the identity matrix. Under the assumption that the saddle point (x^*, λ^*) of $L(x, \lambda) = c^\top x + \sum_{i=1}^n \lambda_i (a_i^\top x - b_i)$ is unique, it can be shown that the trajectories of (19) initialized in $\mathbb{R}^n \times \mathbb{R}^m_{++}$ converge to (x^*, λ^*) and that it is exponentially stable. Thus, the system converges to the solution x^* of (18). The proof can be found in [8].

IV. Example

In the following we show two examples. In the first example we apply the algorithm (5) to the optimization problem

$$\begin{aligned} \min_x \; & (x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{s.t. } & (x_1 - 1)^2 + x_2^2 - 2 \leq 0 \\ & (x_1 + 1)^2 + x_2^2 - 2 \leq 0. \end{aligned} \tag{20}$$

The trajectory of the x-coordinate as well as the feasible set and the level sets of the objective function are shown in Fig. 2. One can also see the trajectory of $\bar{x}(\lambda(t))$ and its tracking by $x(t)$.

We emphasize, that there is no need to choose an initial condition in the feasible set, but only positive initial conditions λ_{i0}.

V. Summary and Future Work

We proposed the vector field (5) whose trajectories converge globally to a saddle point of the Lagrangian function (2) belonging to the constrained optimization problem (1).

Under suitable assumptions such as strict convexity and strict feasibility it is well known that the saddle point of the Lagrangian coincides with the solution of a convex optimization problem. Thus, (5) provides a novel way to solve convex optimization problems of type (1).

One nice feature of system (5) is that the well-known regularity and optimality conditions from optimization theory do not only guarantee well-posedness of the problem but also dictate local and global qualitative behavior of the vector field (5). The main focus of our future research lies in the relaxation of the conditions on the functions f and g_i, $i = 1, \dots, m$, such that we are able to prove global convergence for other classes of optimization problems.

Another interesting research point is to study the underlying geometry of (5). Saddle point problems possess interesting geometric properties (see [6] and [11]). As shown in [6], saddle point algorithms like the AHU-flow have a gradient-like structure using a suitable (indefinite) metric. It is interesting in that respect that the λ-dynamics of (5) can be interpreted as a gradient flow for the dual function with the definite (but singular) metric $\operatorname{diag}[\lambda_1^{-1}, \dots, \lambda_m^{-1}]$.

VI. APPENDIX

A. Derivative of $\bar{x}_c(\lambda)$ with respect to λ

Lemma 5: Suppose (A1) and (A2) are satisfied and let $\bar{x}_c(\lambda)$ denote the unique minimizer of $L_c(x, \lambda)$ for a given $\lambda \in \mathbb{R}^m_+$. Then $\frac{\partial \bar{x}_c(\lambda)}{\partial \lambda} = -\left(\nabla^2 L(\bar{x}_c(\lambda), \lambda)\right)^{-1} \left[\nabla g_1(\bar{x}_c(\lambda)), \dots, \nabla g_m(\bar{x}_c(\lambda))\right]$. By abuse of notation, we write $\nabla^2 L(\bar{x}_c(\lambda), \lambda)$ instead of $\frac{\partial \nabla L(x,\lambda)}{\partial x}|_{\bar{x}_c(\lambda)}$.

Proof: The minimizer $\bar{x}_c(\lambda)$ is implicitly defined as the unique solution of the equation (10). Due to (A1) and (A2), the minimizer exists and is unique. Moreover, due to the Implicit Function Theorem, the derivative of $\bar{x}_c(\lambda)$ with respect to λ exists and can be calculated by differentiating both sides of (10) with respect to λ

$$\frac{\partial \nabla L(x, \lambda)}{\partial x}|_{\bar{x}_c(\lambda)} \frac{\partial \bar{x}_c(\lambda)}{\partial \lambda} + \frac{\partial \nabla L(x, \lambda)}{\partial \lambda}|_{\bar{x}_c(\lambda)} = 0, \quad (21)$$

where $\frac{\partial \nabla L(x,\lambda)}{\partial x}|_{\bar{x}_c(\lambda)}$ coincides with $\nabla^2 L(x, \lambda)|_{\bar{x}_c(\lambda)}$. This yields $\frac{\partial \bar{x}_c(\lambda)}{\partial \lambda} = -\left(\nabla^2 L(\bar{x}_c(\lambda), \lambda)\right)^{-1} \frac{\partial \nabla L(x,\lambda)}{\partial \lambda}|_{\bar{x}_c(\lambda)} = -\left(\nabla^2 L(\bar{x}_c(\lambda), \lambda)\right)^{-1} \left[\nabla g_1(\bar{x}_c(\lambda)), \dots, \nabla g_m(\bar{x}_c(\lambda))\right]$. ■

B. Theorem 6 in [3]

The following assumptions are needed for the Theorem: 1. Consider a manifold $\mathcal{M} \subseteq \mathbb{R}^N$ of class C^2 with standard metric on which a locally Lipschitz continuous vector field $\dot{x} = f(x)$ is given; 2. The solution $x(t)$ is bounded; 3. The ω-limit set $\Omega(x(0))$, which is a compact and connected set, is contained in a closed embedded submanifold $\mathcal{S} \subset \mathcal{M}$. Equivalently, assume that $\mathcal{S}$ is attracting for $x(t)$ starting at $x(0)$; 4. Call O an open tubular neighborhood of $\mathcal{S}$ in $\mathcal{M}$ and assume that there exists a real-valued C^1 height function $W : O \to \mathbb{R}$ and such that $\dot{W} \geq$ on S (or $\dot{W} \leq 0$ on $\mathcal{S}$), where $\dot{W}$ is the derivative of $W(x)$ along the trajectory (Lie derivative). Moreover, let $E := \{x \in \mathcal{S} : \dot{W} = 0\}$ so that $\dot{W} > 0$ on $\mathcal{S} \backslash E$ (or $\dot{W} = 0$ on $\mathcal{S} \backslash E$).

Definition: Let $\{E_i\}_{i \in I}$ be the connected components of E. Given a height function $W(x)$ as in the assumptions above, say that the components $\{E_i\}_{i \in I}$ are *contained* in $W(x)$ if each E_i lies in a level set of $W(x)$, and the subset $\{W(E_i)\}_{i \in I} \subset \mathbb{R}$ has at most a finite number of accumulation points in $\mathbb{R}$.

Theorem 4: Assume the assumptions above hold. If the components $\{E_i\}_{i \in I}$ are contained in $W(x)$ according to the foregoing Definition, then $\Omega(x(0)) \subset E_i$ for a unique $i \in I$.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in Linear and Non-Linear Programming*. Stanford University Press, 1958.

[3] A. Arsie and C. Ebenbauer. Locating omega-limit sets using height functions. *Journal of Differential Equations*, 248(10):2458 – 2469, 2010.

[4] E. N. Barron, R. Goebel, and R. R. Jensen. Best response dynamics for continuous games. *Proc. Amer. Math. Soc.*, 138(3):1069–1083, 2010.

[5] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.

[6] A. Bloch, R. Brockett, and T. Ratiu. On the geometry of saddle point algorithms. In *Proceedings of the 31st IEEE Conference on Decision and Control*, volume 2, pages 1482 –1487, 1992.

[7] G. W. Brown and J. von Neumann. Solutions of games by differential equations. 1:73–79, 1950.

[8] H.-B. Dürr, S.-S. Zeng, and C. Ebenbauer. Ein nichtlineares System zum Lösen von Linearen Programmen und Sattelpunktproblemen. *17. Steirisches Seminar über Regelungstechnik und Prozessautomatisierung, Graz, Austria*, 2011.

[9] K. H. Elster. *Nichtlineare Optimierung*. Harri Deutsch, 1978.

[10] D. Feijer and F. Paganini. Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica*, 46(12):1974–1981, 2010.

[11] O. Fujiwara. Morse Programs: A topological approach to smooth constrained optimization. *Mathematics of Operations Research*, 7(4):pp. 602–616, 1982.

[12] C. Geiger and C. Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, 2002.

[13] U. Helmke and J. B. Moore. *Optimization and dynamical systems*. Springer, 1994.

[14] J. Hofbauer. Best response dynamics for continuous zero–sum games best response dynamics for continuous zero–sum games best response dynamics for continuous zero–sum games. *Discrete and Continuous Dynamical Systems*, 6(1):215–224, 2006.

[15] A. Iggidr and G. S. Sallet. On the stability of nonautonomous systems. *Automatica*, 39(1), 2003.

[16] H. Khalil. *Nonlinear Systems*. Prentice Hall, 2002.

[17] T. Kose. Solutions of saddle value problems by differential equations. *Econometrica*, 24(1):pp. 59–70, 1956.

[18] S. Krantz and H. R. Parks. *A Primer of Real Analytic Functions*. Birkhäuser, 2002.

[19] Natanson. *Theory of Functions of a Real Variable*. Ungar Pub Co, 1961.

[20] A. Nedić and A. Ozdaglar. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1):205–228, 2009.

[21] R. T. Rockafellar. *Convex Analysis (Princeton Mathematical Series)*. Princeton University Press, 1970.

[22] H. Whitney. A function not constant on a connected set of critical points. *Duke Mathematical Journal*, 1935.