# Projected Gradient Descent on Riemannian Manifolds with Applications to Online Power System Optimization

Adrian Hauswirth[1,2], Saverio Bolognani[1], Gabriela Hug[2], and Florian Dörfler[1]

***Abstract*— Motivated by online optimization problems arising in nonlinear power system applications, this article concerns optimization over closed subsets of Riemannian manifolds.**

**Compared to conventional optimization over manifolds, we explicitly consider inequality constraints that result in a feasible set that is itself not a smooth manifold. We propose a continuous-time projected gradient descent algorithm over the feasible set and show its well-behaved convergent behavior. Under mild assumptions on the non-degeneracy of equilibria we show that points are local minimizers if and only if they are asymptotically stable.**

**The proposed algorithm can be implemented as a real-time feedback control law on a physical system. This approach is particularly appropriate for online load flow optimization problem in power systems, in which the state of the grid is naturally constrained to the manifold that represents the solution space to the nonlinear AC power flow equations.**

**We specialize our approach for the case of power distribution systems that need to respect operational constraints while being economically efficient, and we illustrate the resulting closed-loop behavior in simulations.**

## I. Introduction

Finding the minimal value and minimizer of a function on a given closed domain is one of the most fundamental problems in mathematics and arises in many engineering problems. One of the most studied problem class is the optimization of a function over a subset of Euclidean space. The methods applied to these problems can often be defined on spaces that are only locally Euclidean, namely differentiable manifolds. Given the vast variety of differentiable manifolds and engineering problems defined on these manifolds, this extension considerably increases the scope of traditional concepts for optimization. A differential geometric approach has led to elegant reformulations of problems like eigenvalue computations and sorting [1] that can be cast as flows on manifolds. More recently, it has fueled the development of practical algorithms for numerical computations [2], [3].

To the best of the authors' knowledge, all of these works have only considered unconstrained manifold optimization problems where the only constraint is the fact that the solution has to lie on the manifold. However, this setup severely limits the scope of possible applications since many problems involve not only manifold-constraints, but also additional restrictions on the space of feasible points, particularly in engineering problems. These will often take the form of inequality constraints.

This research is supported by ETH start-up funds and the SNF Assistant Professor Energy Grant #160573.

[1] Automatic Control Laboratory, ETH Zürich, Switzerland.
[2] Power Systems Laboratory, ETH Zürich, Switzerland.
Email: {hadrian,bsaverio,hug,dorfler}@ethz.ch

In this paper we consider constrained optimization over manifolds, that is, the optimization of a function over a subset of a manifold. In analogy to the Euclidean case, we define a continuous-time projected gradient descent algorithm. We show that its trajectories asymptotically converge to equilibrium points, and that under suitable non-degeneracy assumptions these equilibria are stable if and only if they are local minimizers.

We show how this type of algorithm can be realized in a controlled physical differential-algebraic system that is drift-free, i.e., remain stationary in the absence of control inputs. This allows us to use the proposed approach for online optimization in closed-loop. Instead of solving the optimization problem offline and applying the final solution to the system, each iterate of the optimization is fed back to the system with the effect of being more computationally efficient, robust towards parameter uncertainty, and potentially amenable to distributed implementation.

Online optimization has recently become popular in the field of power system, for voltage regulation problems [4], optimal reactive power compensation [5], optimal frequency control [6] and economic dispatch [7], [8]. We adopt the same application, and in particular load flow optimization problem in a power distribution grid, to illustrate the potential of the approach proposed in this paper.

The rest of this paper is structured as follows. Section II summarizes various results from convex and variational analysis, differential and Riemannian geometry and key concepts for discontinuous dynamical systems. In Section III we define a class of constrained optimization problems on manifolds and introduce projected gradient descent as a continuous-time solution algorithm. We show its asymptotic convergence to equilibrium points and relate their stability to local optimality.

Section IV shows how projected gradient descent can be locally implemented as an online feedback control law on a given drift-free physical system. Finally, in Section V we simulate its behavior in a simplified power system.

### A. Notation

Given a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$, we write $\nabla f(x) \in \mathbb{R}^{m\times n}$ for the Jacobian matrix in standard coordinates. The kernel space of $\nabla f(x)$ as linear map is denoted by $\ker \nabla f(x)$.

For a differentiable curve $\gamma : \mathbb{R} \to \mathbb{R}^n$ we use the notation $\gamma'(t) = d\gamma/dt(\tau)$ and $\gamma'_+(\tau) = \lim_{t\to\tau^+} \gamma'(t)$.

## II. Preliminaries

### A. Projection on Closed Convex Cones

We recall some basic facts about projections on cones. Let $V$ be a finite-dimensional vector space, then $C \subseteq V$ is a cone if for all $x \in C$ it holds that $\lambda x \in C$ for all $\lambda \geq 0$. The cone $C$ is closed and convex if it is closed and convex as a subset of $V$. Consider an inner product $\langle \cdot, \cdot \rangle_r : V \times V \to \mathbb{R}$, that is a positive-definite, symmetric bilinear form on $V$. The dual cone $C^*$ of $C$ is defined as $C^* := \{v | \langle v, w \rangle_r \geq 0 \forall w \in C\}$.[1] Let $||\cdot||_r$ denote the 2-norm induced by $\langle \cdot, \cdot \rangle_r$. Given $x \in V$, its projection on $C$ is defined as $\Pi^r_C(x) := \arg\min_{y \in C} ||x - y||^2_r$. If $C$ is closed and convex, the projection exists and is unique. Furthermore, the following properties hold.

**Lemma 1.** *[9, Proposition 3.2.3] Given a closed convex cone $C \subset V$, then $\tilde{x}$ is the projection of $x \in V$ if and only if there exists $n \in C^*$ such that*

$$\tilde{x} = x + n \quad \text{and} \quad \langle n, \tilde{x} \rangle_r = 0 \ .$$

Furthermore, the projection on a closed convex cone is idempotent and non-expansive [10, E.9.3], that is,

$$||\Pi^r_C(x_1) - \Pi^r_C(x_2)||_r \leq ||x_1 - x_2||_r \ .$$

In the special case where $V = \mathbb{R}^n$ is endowed with the canonical basis and with the Euclidean norm $e$, and $C$ is a subspace, then the projection on $C$ takes an explicit form which will be useful later.

**Lemma 2.** *Let $A \in \mathbb{R}^{m \times n}$ have rank $m \leq n$. The projection of $x \in \mathbb{R}^n$ on $C = \ker A$ is given as*

$$\Pi^e_{\ker A}(x) = x - A^T \beta \ .$$

*where $\beta := (AA^T)^{-1} Ax$ .*

*Proof.* First, note that the dual cone of $\ker A$ is the space spanned by the columns of $A^T$. In accordance with Lemma 1, we define $n := -A^T \beta$ and therefore just need to check that $\langle n, \tilde{x} \rangle = 0$, which can be verified by inspection. □

### B. Inward tangent and normal cones

To properly define a projected gradient descent we need to describe the set of admissible directions at every point of a set. In the interior of the feasible set, any direction is admissible. At the boundary however, a vector represents an admissible direction only if it points into the set. These concepts are made formal using the following notions from [11].

Given a set $\mathcal{K} \subset \mathbb{R}^n$, a vector $v \in \mathbb{R}^n$ is called a *(geometrically derivable) inward tangent vector at $x \in \mathcal{K}$* if there exists a smooth curve $\gamma : [0, \epsilon) \to \mathcal{K}$ with $\epsilon \geq 0$, $\gamma(0) = x$ and $\gamma'_+(0) = v$. We call the set of all inward tangent vectors at $x$ the *inward tangent cone* and denote it by $T^>_x \mathcal{K}$.

It is easy to see that $T^>_x \mathcal{K}$ is indeed a cone, since any curve $t \mapsto \gamma(t)$ with $\gamma'_+(0) = v$ can be reparametrized as $t \mapsto \tilde{\gamma}(t)$ where $\tilde{\gamma}(t) := \gamma(\lambda t)$ for $\lambda > 0$. The chain rule yields $\tilde{\gamma}'_+(0) = \lambda v$.

Given $\mathcal{K}$ with inward tangent cone $T^>_x \mathcal{K}$ at $x$, the *inward normal cone* is defined as the dual cone of $T^>_x \mathcal{K}$, that is,

$$N^>_x \mathcal{K} := \{w \in \mathbb{R}^n | \langle w, v \rangle \geq 0 \, \forall v \in T^>_x \mathcal{K}\} \ .$$

In the rest of the paper we will often consider sets that are of the form

$$\mathcal{K} = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\} \tag{1}$$

with $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ smooth.

**Assumption 1.** *Given a set $\mathcal{K}$ as in* (1) *assume that for every $x \in \mathcal{K}$ the vectors $\nabla h_i(x)$ and $\nabla g_j(x)$ are linearly independent for all $i = 1, \ldots, m$ and $j = 1, \ldots, p$ such that $g_j(x) = 0$.*

*Remark* 1. It can be shown that Assumption 1 holds generically [12]. That means, if $g$ and $h$ do not satisfy Assumption 1, an infinitesimally small perturbation of the constraints is sufficient to make the assumption hold.

*Remark* 2. Assumption 1 is also known as LICQ (linear independence constraint qualification) in nonlinear programming.

As a consequence of Theorems 6.30, 6.31 and 6.42 of [11], under Assumption 1 the inward tangent cone $T^>_x \mathcal{K}$ at every $x \in \mathcal{K}$ can be decomposed as follows.

**Theorem 1.** *Let $\mathcal{K}$ be as in* (1) *and satisfy Assumption 1. Define $\mathcal{H}_i := \{x | h_i(x) = 0\}$ and $\mathcal{G}_j := \{x | g_j(x) \leq 0\}$ where $h_i$ and $g_j$ denote the $i$-th and $j$-th component function of $h$ and $g$ respectively. Then, the inward tangent cones of $\mathcal{H}_i$ and $\mathcal{G}_j$ at $x$ are given as*

$$T^>_x \mathcal{H}_i = \{v \in \mathbb{R}^n | \nabla h_i(x) v = 0\}$$
$$T^>_x \mathcal{G}_j = \{v \in \mathbb{R}^n | \nabla g_j(x) v \leq 0\} \ .$$

*Furthermore, the inward tangent cone of $\mathcal{K}$ is closed convex and is given by*

$$T^>_x \mathcal{K} = \left( \bigcap_{i=1}^{m} T^>_x \mathcal{H}_i \right) \cap \left( \bigcap_{j=1}^{p} T^>_x \mathcal{G}_j \right)$$

### C. Differential and Riemannian geometry

In the following, we introduce a few basic notions from differential geometry required for our particular application. For a comprehensive introduction see [13], [14].

We consider regular submanifolds of $\mathbb{R}^n$ of the form

$$\mathcal{M} = \{x \in \mathbb{R}^n | h(x) = 0\} \tag{2}$$

where $h : \mathbb{R}^n \to \mathbb{R}^q$ is a smooth function such that for every $x \in h^{-1}(0)$ the rank of $\nabla h(x)$ is $q$. Then, $\mathcal{M}$ is called a *smooth (embedded) manifold of dimension $n - q$*.

To every point $x \in \mathcal{M}$ there is a *tangent space* $T_x \mathcal{M}$ attached, identified as

$$T_x \mathcal{M} = \ker \nabla h(x) = T^>_x \mathcal{M} \ . \tag{3}$$

The second equality emphasizes the relation to the overlapping, yet consistent, definitions of the previous section.

[1] Strictly speaking, the dual cone $C^*$ is a cone in the dual space of $V$, but we may as well consider $C^*$ to be a subset of $V$ under the isomorphism between $V$ and its dual implicitly defined by the inner product.

Namely, for a smooth manifold its tangent space (as a manifold) is identical to its inward tangent cone (as a set).

For a smooth function $\phi : \mathcal{M} \to \mathbb{R}$, and $x \in \mathcal{M}$, the *differential* $D_x\phi : T_x\mathcal{M} \to \mathbb{R}$ is defined as

$$D_x\phi(v) := (\phi \circ \gamma)'(v)$$

where $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ is a smooth curve with $\gamma(0) = x$ and $\gamma'(0) = v \in T_x\mathcal{M}$. It can be shown that the differential is independent of the curve $\gamma$. As such, the differential can be interpreted as a directional derivative in direction $v$.

A *vector field* $F$ is map that assigns to every point $x \in \mathcal{M}$ a vector $F(x) \in T_x\mathcal{M}$.

Given the smooth manifold $\mathcal{M}$, a *Riemannian metric* $r$ on $\mathcal{M}$ is a smooth map that assigns to every point $x \in \mathcal{M}$ an inner product $r(x) : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$ denoted by $\langle u, v\rangle_r$ for $u, v \in T_x\mathcal{M}$. Besides being the foundation of many advanced concepts, a Riemannian metric $r$ is required to define the notion of gradient on a manifold. Given a function $\phi : \mathcal{M} \to \mathbb{R}$, the *gradient of* $\phi$ *at* $x \in \mathcal{M}$ is defined as the unique tangent vector $\operatorname{grad}\phi(x)$ such that

$$\langle \operatorname{grad}\phi(x), v\rangle_r = D_x\phi(v) \tag{4}$$

holds for every $v \in T_x\mathcal{M}$.

It is a well known fact that, the gradient of $f$ at $x$ is the direction of *steepest ascent* in the sense that

$$\frac{\operatorname{grad}\phi(x)}{||\operatorname{grad}\phi(x)||_r} = \arg \max_{v\in T_x\mathcal{M}:\, ||v||_r=1} D_x\phi(v) \ .$$

Since $\mathcal{M}$ is embedded in $\mathbb{R}^n$ let $f : \mathbb{R}^n \to \mathbb{R}$ be such that $\phi = f|_\mathcal{M}$ and by choosing standard coordinates we can write $D_x\phi(v) = \nabla f(x)v$ for $v \in T_x\mathcal{M}$.

In the case study in Section V we will use the the metric induced by the ambient Euclidean space given by

$$\langle\cdot,\cdot\rangle_r : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}, \qquad (v, w) \mapsto \langle v, w\rangle_e \tag{5}$$

where $\langle\cdot,\cdot\rangle_e$ denotes the standard inner product on $\mathbb{R}^n$.

With this induced metric the definition of the gradient on $\mathcal{M}$ reduces to a projection of $\nabla f(x)$ on $T_x\mathcal{M}$, that is,

$$\operatorname{grad}\phi(x) = \Pi^e_{\ker \nabla h(x)}(\nabla f(x)^T) \ , \tag{6}$$

as one can verify using the definition for $\Pi^e_{\ker \nabla h(x)}$ provided in Lemma 2.

### *D. Discontinuous dynamical systems*

Consider a dynamical system $\dot{x} = F(x)$ where $F$ is a vector field defined on a manifold $\mathcal{M}$.

Given $x_0 \in \mathcal{M}$, a curve $\gamma : [0, \epsilon) \to \mathcal{M}$ for some $\epsilon > 0$ is a called a *(Carathéodory) solution* of $\dot{x} = F(x)$ on the interval $[0, T)$ if it is absolutely continuous[2], $\gamma(0) = x_0$ and if $\dot{x}(t) = F(x(t))$ holds except on a zero measure set.

A set $\mathcal{S} \subseteq \mathcal{M}$ is called *invariant* if every solution starting at some $x_0 \in \mathcal{S}$ remains in $\mathcal{S}$.

A point $x^* \in \mathcal{M}$ is called an *equilibrium point of* $F$ if $F(x^*) = 0$. An equilibrium point $x^*$ is called *isolated* if there exists an open neighborhood $\mathcal{U}$ such that $\mathcal{U} \setminus \{x^*\}$ contains no equilibrium point.

An equilibrium point $x^*$ is called *(Lyapunov) stable* if for every neighborhood $\mathcal{U}$ of $x^*$ there exists another neighborhood $\mathcal{V} \subset \mathcal{U}$ of $x^*$ such that for every solution $x$ with $x(0) \in \mathcal{V}$ it holds that $x(t) \in \mathcal{U}$ for $t \geq 0$. An equilibrium point $x^*$ is *asymptotically stable* if it is stable and there exists a neighborhood $\mathcal{U}$ of $x^*$ such that all trajectories starting in $\mathcal{U}$ converge to $x^*$.

Given a smooth function $\phi : \mathcal{M} \to \mathbb{R}$, the *Lie derivative of* $\phi$ along the vector field $F(x)$ is given by $\mathcal{L}_F\phi(x) := D_x\phi(F(x))$. If $\mathcal{M}$ is embedded in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is such that $\phi = f|_\mathcal{M}$ then $\mathcal{L}_F\phi(x) = \nabla f(x)F(x)$ for all $x \in \mathcal{M}$ assuming canonical coordinates.

For the convergence analysis we require a LaSalle-like invariance theorem for discontinuous systems that goes back to [16]. We use the following simplified version.

**Theorem 2.** *[17, Proposition 2.1] Let $F$ be a vector field on a domain $\mathcal{D} \subset \mathbb{R}^n$ and let $\mathcal{S} \subset \mathcal{D}$ be compact and invariant such that for every $x_0 \in \mathcal{S}$ there is a unique global solution. Let $f : \mathcal{S} \to \mathbb{R}$ be continuously differentiable such that $\mathcal{L}_F f(x) \leq 0$ for all $x \in \mathcal{S}$. Then, the Carathéodory solution $x$ starting at $x_0 \in \mathcal{S}$ converges to the largest invariant set in the closure of*

$$\{x \in \mathcal{S} | \mathcal{L}_F f(x) = 0\} \ .$$

## III. Constrained Optimization over Manifolds

We now turn to the main problem of finding the minimum (and minimizer) of a function on a subset of a manifold.

To the authors' knowledge, research in the field of optimization on manifolds has almost exclusively considered "unconstrained" problems in the sense that a solution can lie anywhere on the manifold but is not otherwise constrained.

However, for practical problems, such as power systems optimization, it is necessary to consider optimization over feasible subsets describing operational constraints.

Hence, a constrained optimization problem over the manifold can be defined as

$$\begin{aligned} \underset{x\in\mathcal{M}}{\text{minimize}} \quad & \phi(x) \\ \text{subject to} \quad & x \in \mathcal{K} \end{aligned} \tag{7}$$

for a smooth $\phi : \mathcal{M} \to \mathbb{R}$ and a closed subset $\mathcal{K}$ of $\mathcal{M}$.

A point $x^* \in \mathcal{K}$ is a *minimizer* of (7) if there exists a neighborhood $\mathcal{U} \subseteq \mathcal{K}$ of $x^*$ such that $\phi(x^*) \leq \phi(x)$ for all $x \in \mathcal{U}$. A minimizer $x^*$ is *strict* if there exists a neighborhood $\mathcal{U}$ of $x^*$ such that $\phi(x^*) < \phi(x)$ for all $x \in \mathcal{U} \setminus \{x^*\}$.

For our control application we restrict ourselves to submanifolds of $\mathbb{R}^n$. Recall, that for our purposes $\mathcal{M}$ takes the form (2), i.e., it is the regular level set of some function $h : \mathbb{R}^n \to \mathbb{R}^q$ and is endowed with a Riemannian metric $r$. Furthermore, we assume that there exist smooth functions $f : \mathbb{R}^n \to \mathbb{R}$ such $\phi := f|_\mathcal{M}$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ such that

$$\mathcal{K} = \mathcal{M} \cap \mathcal{G} \tag{8}$$

[2] A curve on a smooth manifold is absolutely continuous if it is absolutely continuous in any coordinate domain. This definition goes back to [15].

where $\mathcal{G} := \{x \in \mathbb{R}^n | g(x) \leq 0\}$. Then, (7) has the same solution as the nonlinear programming problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \\ & g(x) \leq 0 \ . \end{aligned}$$

However, in the spirit of optimization over manifolds, we are only interested in algorithms that produce iterates or trajectories that lie on $\mathcal{K}$ at all times. This requirement is particularly important when considering online optimization in closed loop where instead of applying the final solution to a system, the iterates or trajectory obtained during runtime are fed into the system. For this to work properly, the iterates have to be feasible at all times.

We proceed by defining a continuous-time algorithm for solving (7) that takes the form of a (discontinuous) dynamical system on $\mathcal{M}$. Our algorithm is inspired by the theory on projected dynamical systems over convex sets in $\mathbb{R}^n$ [18].

### A. Projected dynamical system on Riemannian manifolds

Among the most basic optimization paradigms is *gradient descent*. By moving in the along direction of the steepest descent of a differentiable function, one is guaranteed to asymptotically reach a local minimum (unless the problem is unbounded or the initial point is a equilibrium that is not a minimizer). This is straightforward for unconstrained optimization. In the case of constrained optimization, this approach is complicated by the fact that at the boundary of the feasible set the gradient can point in an unfeasible direction, i.e., pointing out of the feasible set. The solution to this is to restrict the set of admissible directions and find the direction of steepest descent within the inward tangent cone. To accomplish this, we define the following operator that projects any tangent vector on the inward tangent cone.

**Definition 1.** *For a Riemannian manifold $(\mathcal{M}, r)$ and a closed set $\mathcal{K} \subset \mathcal{M}$ for which $T_x^{>}\mathcal{K}$ is a closed convex cone define the operator*

$$\begin{aligned} \text{proj}_{\mathcal{K}}(x, \cdot): \quad T_x\mathcal{M} &\rightarrow T_x^{>}\mathcal{K} \\ v &\mapsto \Pi^r_{T_x^{>}\mathcal{K}}(v) \end{aligned}$$

*for every $x \in \mathcal{K}$.*

Recall that $\mathcal{M}$ is a submanifold of $\mathbb{R}^n$ such that $\mathcal{K}$ takes the form (8). Under Assumption 1 we can apply Theorem 1 and conclude that $T_x\mathcal{K}$ is closed convex for all $x \in \mathcal{K}$. Furthermore, using the induced metric (5) on $\mathcal{M}$ and a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $f|_{\mathcal{M}} = \phi$, the evaluation of $\text{proj}_{\mathcal{K}}(x, -\operatorname{grad}\phi(x))$ reduces to solving a quadratic program

$$\begin{aligned} \underset{w\in\mathbb{R}^n}{\text{minimize}} \quad & ||\nabla f(x) - w||^2 \\ \text{subject to} \quad & \nabla h(x) w = 0 \\ & \nabla g_{J(x)}(x) w \leq 0 \end{aligned} \tag{9}$$

where $J(x) = \{i | g_i(x) = 0\}$ is the index set of active constraints at $x$.

Given a smooth function $\phi : \mathcal{K} \rightarrow \mathbb{R}$, we can apply $\text{proj}_{\mathcal{K}}$ to its gradient to yield the *projected gradient vector field* $\text{proj}_{\mathcal{K}}(x, -\operatorname{grad}\phi(x))$ for all $x \in \mathcal{K}$.

For the rest of the paper we will consider the *projected gradient descent algorithm* given by the dynamical system

$$\dot{x} = \text{proj}_{\mathcal{K}}(x, -\operatorname{grad}\phi(x)) \qquad x(0) = x_0 \tag{10}$$

where $x_0 \in \mathcal{K}$ is an initial value.

In general, as (10) has a discontinuous right-hand side, standard results about the existence and uniqueness of solution trajectories do not apply.

**Assumption 2.** *Given a set $\mathcal{K}$ of the form* (8) *satisfying Assumption 1, assume that for every $x_0 \in \mathcal{K}$ there exists a unique solution $x(t)$ in the sense of Carathéodory to* (10) *on the interval $[0, \infty)$ such that $x(0) = x_0$. Furthermore, $x(t) \in \mathcal{K}$ for all $x \in [0, \infty)$.*

*Remark* 3. It is tempting to think that Assumption 2 is guaranteed by the existence and uniqueness results of [19] for Euclidean spaces since $\mathcal{M}$ is locally Euclidean. However, upon inspection, projected systems on manifolds fall into the category of projected dynamical systems with *oblique projections* – a case not covered in [19]. Nevertheless, under Assumption 1, the results from [20], that constitute the key ingredient in the proofs in [19], apply to this more general case with oblique projections, and it is then plausible to assume that the results generalize to Assumption 2.

### B. Convergence and Local Optimality of Projected Gradient Descent

We now show that projected gradient descent asymptotically converges to an equilibrium point and that therefore it does not exhibit limit cycles or strange attractors. Additionally, the stability of equilibrium points can be related to their local optimality.

We first show that $\phi$ is non-increasing along trajectories of the projected gradient flow.

**Lemma 3.** *Let $F(x) := \text{proj}_{\mathcal{K}}(x, -\operatorname{grad}\phi(x))$ be the projected gradient field with respect to the metric $r$. Then*

$$\mathcal{L}_F\phi(x) \leq 0 \qquad \forall x \in \mathcal{K}.$$

*Proof.* Using Lemma 1 we know that

$$\text{proj}_{\mathcal{K}}(x, -\operatorname{grad}\phi(x)) = -\operatorname{grad}\phi(x) + n(x)$$

for some $n(x) \in N_x^{>}\mathcal{K}$. Hence, using the definition of the gradient in (4) we can write

$$\begin{aligned} \mathcal{L}_F\phi(x) &= -D_x\phi(\operatorname{grad}\phi(x)) + D_x\phi(n(x)) \\ &= -||\operatorname{grad}\phi(x)||_r^2 + \langle \operatorname{grad}\phi(x), n(x)\rangle_r \ . \end{aligned}$$

We therefore need to show that $\langle \operatorname{grad}\phi(x), n(x)\rangle_r \leq ||\operatorname{grad}\phi(x)||_r^2$. In particular, using the Cauchy-Schwarz inequality on $T_x\mathcal{M}$ with inner product $r(x)$, it suffices to show that $||n(x)||_r \leq ||\operatorname{grad}\phi(x)||_r$. This in turn follows directly from the Pythagorean theorem. Consequently, we have $\mathcal{L}_F\phi(x) \leq 0$. □

Based on Lemma 3, the next lemma shows that for projected gradient descent sublevel sets of the potential function are invariant.

**Lemma 4.** *Under Assumption 2, the sublevel set $\mathcal{S}_p := \{x \in \mathcal{K} | \phi(x) \leq p\}$ is invariant with respect to the projected gradient descent* (10)*.*

*Proof.* Assume for the sake of contradiction that there exists a Carathéodory solution $x(t)$ starting at $x_0 \in \mathcal{S}_p$ and at $t = T$ we have $x(T) \notin \mathcal{S}_p$. Since $x$ is absolutely continuous and $\phi$ is smooth we conclude that $(\phi \circ x)(t)$ is absolutely continuous. In particular, $(\phi \circ x)(t)$ has a derivative almost everywhere given by $(\phi \circ x)'(t) = \mathcal{L}_F \phi(x)$ such that

$$(\phi \circ x)(T) = (\phi \circ x)(0) + \int_0^T \mathcal{L}_F \phi(x(t)) dt \ .$$

However, due to Lemma 3 $\mathcal{L}_F \phi(x) \leq 0$ for all $x \in \mathcal{K}$ and therefore we have $\int_0^T \mathcal{L}_F \phi(x(t)) dt \leq 0$ which leads to the contradiction. □

**Theorem 3.** *Assume that $\phi$ has compact level sets and that Assumption 2 holds for projected gradient descent. Then, all Carathéodory trajectories to $\dot{x} = \operatorname{proj}_{\mathcal{K}}(x, -\operatorname{grad} \phi)$ converge asymptotically to the set of equilibrium points.*

*Proof.* Since $\mathcal{K} \subset \mathbb{R}^n$ consider $\operatorname{proj}_{\mathcal{K}}(x, -\operatorname{grad} \phi(x)) : \mathcal{K} \to \mathbb{R}^n$ to be a vector field in $\mathbb{R}^n$.

According to Lemma 4 the sublevel sets $\mathcal{S}_p \subset \mathcal{K}$ of $\phi$ are invariant with respect to the projected gradient descent and by assumption they are compact. Consequently, Theorem 2 applies and any trajectory starting in some sublevel set $\mathcal{S}_p$ converges to the largest invariant set in the closure of $\{x \in \mathcal{S}_p | \mathcal{L}_F \phi(x) = 0\}$.

Finally, we show that the projected gradient asymptotically converges to an equilibrium point. To see this, note that according to the proof of Lemma 3, $\mathcal{L}_F \phi(x) = 0$ implies that either $\operatorname{grad} \phi(x) = 0$ or $n(x) = \operatorname{grad} \phi(x)$. In both cases we have, that

$$\operatorname{proj}_{\mathcal{K}}(x, -\operatorname{grad} \phi(x)) = 0$$

and hence $x$ is an equilibrium point. □

It is not a priori clear whether an equilibrium $x$ of $\operatorname{proj}_{\mathcal{K}}(x, -\operatorname{grad} \phi(x))$ is a local minimum of $\phi$ in $\mathcal{K}$. Interestingly, the stability of equilibria to can be related to their optimality. This is particularly useful in view of online closed-loop optimization.

First, note the following easy result.

**Lemma 5.** *Let $x^*$ be a local minimizer of $\phi$ on $\mathcal{K}$. Then $x^*$ is an equilibrium point of the projected gradient vector field* $\operatorname{proj}_{\mathcal{K}}(x, -\operatorname{grad} \phi(x))$.

*Proof.* The proof is similar to the proof of Lemma 4. □

**Theorem 4.** *Assume that $\phi$ has compact level sets and that Assumption 2 holds for projected gradient descent. Then the following are true:*

(1) *If $x^* \in \mathcal{K}$ is an asymptotically stable equilibrium point of* (10) *then it is a strict local minimum of $\phi$ on $\mathcal{K}$.*
(2) *If $x^* \in \mathcal{K}$ is a strict local minimum of $\phi$ on $\mathcal{K}$ then it is a stable equilibrium point of* (10)*.*

*Proof.* For (1), let $\mathcal{U}$ be a neighborhood of $x^*$ such any solution to (10) with $x(0) \in \mathcal{U}$ converges to $x^*$. As in the proof of Lemma 4 we may write

$$\lim_{t \to +\infty} (\phi \circ x)(t) = \phi(x^*) = (\phi \circ x)(0) + \int_0^{+\infty} \mathcal{L}_F \phi(x(t)) dt \ .$$

However, since $\int_0^{+\infty} \mathcal{L}_F \phi(x(t)) \leq 0$ it follows that $\phi(x^*) \leq \phi(x(0))$ and therefore, $x^*$ is a local minimizer of $\phi$. To see that it is a strict minimizer, assume for the sake of contradiction that for some $\tilde{x}$ in the region of attraction $\mathcal{U}$ of $x^*$ it holds that $\phi(\tilde{x}) = \phi(x^*)$. The solution starting at $\tilde{x}$ nevertheless converges to $x^*$ by assumption. Therefore, it must hold that $\int_0^{+\infty} \mathcal{L}_F \phi(x(t)) = 0$ and since $\mathcal{L}_F \phi \leq 0$ it follows that $\mathcal{L}_F \phi(x(t)) = 0$ for almost all $t \geq 0$. But as a consequence of Theorem 3 all points $x$ with $\mathcal{L}_F \phi(x) = 0$ are equilibrium points, and consequently $x^*$ cannot be asymptotically stable for the neighborhood $\mathcal{U}$.

To show (2), we proceed similarly to the proof of Lyapunov's Theorem [21]. Given any neighborhood $\tilde{\mathcal{U}}$ of $x^*$ let $\mathcal{U} \subseteq \tilde{\mathcal{U}}$ be a compact neighborhood of $x^*$ in which $x^*$ is a strict minimum.

Next, we construct a neighborhood $\mathcal{V} \subset \mathcal{U}$ such that all trajectories starting $\mathcal{V}$ remain in $\mathcal{U}$ and therefore in $\tilde{\mathcal{U}}$, thus establishing stability.

Let $\alpha$ be such that $\phi(x^*) < \alpha < \min_{x \in \partial \mathcal{U}} \phi(x)$ where $\partial \mathcal{U}$ is the boundary of $\mathcal{U}$. Define $\mathcal{V} := \{x \in \mathcal{U}, \phi(x) \leq \alpha\} \subseteq \mathcal{U}$ which is has a non-empty interior because $\phi(x^*) < \alpha$. We claim that trajectories starting in $\mathcal{V}$ remain in $\mathcal{U}$.

To see this, assume for the sake of contradiction that there is a solution $x$ with $x(0) \in \mathcal{U}$ and $x(T) \notin \mathcal{U}$ for some $T > 0$. By continuity of $x$ there exists $\tau \in (0, T)$ and $x(\tau) \in \partial \mathcal{U}$. That is, at $t = \tau$ the trajectory crosses the boundary of $\mathcal{U}$. However, $\phi(\tau) > \alpha$ by definition of $\alpha$, and therefore $x(\tau) \notin \mathcal{S}_\alpha$ which contradicts Lemma 4. Consequently, solutions starting in $\mathcal{V}$ remain in $\mathcal{U} \subseteq \tilde{\mathcal{U}}$, completing the proof that $x^*$ is stable. □

*Remark* 4. It seems plausible that strict minimizers are asymptotically stable. This, however, is not true in general as the counter-example in [22] shows. Similarly, minimizers are not guaranteed to be stable and stable equilibria are not in general minimizers.

If the equilibrium points of $\phi$ on $\mathcal{K}$ are assumed to be isolated, Theorem 4 can be strengthened using the following result.

**Lemma 6.** *Assume that all equilibrium points of* (10) *are isolated. Then, every stable equilibrium is asymptotically stable.*

*Proof.* This can be shown using Theorem 3. □

From Lemmas 5 and 6 and Theorem 4 it follows that:

**Corollary 1.** *If all equilibrium points of* (10) *are isolated then every minimizer is strict and every stable equilibrium is asymptotically stable. Furthermore, $x^*$ is a (strict) minimizer of $\phi$ on $\mathcal{K}$ if and only if it is an (asymptotically) stable equilibrium.*

## IV. Closed-Loop Online Optimization

In this section we show how we can implement projected gradient descent for a system that is naturally constrained to a subset of a manifold. In practice, such a subset can either be given by hard physical constraints on the system or it can be an attractive set of equilibrium points of an underlying nonlinear dynamical system that evolves on a faster time-scale. It can also be a combination of the two. However, we assume that such as system is *drift-free*, i.e., if the control input is zero the system remains stationary [23].

Let $\mathcal{M} \subset \mathbb{R}^{n_1+n_2}$ be a smooth $n_1$-manifold given as $\mathcal{M} := \{x | h(x) = 0\}$ for $h : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^{n_2}$ smooth with 0 as a regular value. Let $r$ be a Riemannian metric on $\mathcal{M}$ and define $x = \begin{bmatrix} x_1^T & x_2^T \end{bmatrix}^T$ where $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$.

Given a physical system with configuration space $\mathcal{M}$, assume that the variables $x_1$ are *exogenous*, i.e., control variables. The variables $x_2$ are *endogenous*, that is, their value is determined solely by $x_1$. Although, given $x_1$, the value of $x_2$ is not necessarily unique.

Let $u : \mathcal{M} \to \mathbb{R}^{n_1}$ be a control law, $x_0 \in \mathcal{M}$ and consider the differential-algebraic, drift-free system given by

$$\begin{aligned} \dot{x}_1 &= u(x) \\ 0 &= h(x) \end{aligned} \tag{11}$$

with initial condition $x(0) = x_0$ such that $h(x_0) = 0$. We show that by choosing $u$ appropriately, the system (11) behaves locally like a projected gradient descent (10).

For the following lemma define the projection onto the first $n_1$ components as $\pi_1 : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^{n_1}$ such that $\pi_1(x) := x_1$.

**Lemma 7.** *Let $x \in \mathcal{M}$ be such that $\operatorname{rank} \nabla_{x_2} h(x) = m$. Then, for every $v_1 \in \mathbb{R}^m$ there is a unique $v \in T_x\mathcal{M}$ such that $\pi_1(v) = v_1$ and which is given by*

$$v := \begin{bmatrix} v_1 \\ -\left(\nabla_{x_2} h(x)\right)^{-1} \nabla_{x_1} h(x) v_1 \end{bmatrix} \tag{12}$$

*Proof.* For $x \in \mathcal{M}$ such that $\operatorname{rank} \nabla_{x_2} h(x) = m$ the implicit function theorem implies that there is a smooth function $v : U \to V$ where $U \subset \mathbb{R}^{n_1}$ and $V \subset \mathbb{R}^{n_2}$ are open sets such that $x_1 \in U$ and $v(y) \in V$ and $h(y, v(y)) = 0$ for all $y \in U$. Furthermore, we have

$$\nabla v(y) = -\left(\nabla_{x_2} h(y, v(y))\right)^{-1} \nabla_{x_1} h(y, v(y))$$

and in particular, it holds that

$$\begin{bmatrix} \nabla_{x_1} h & \nabla_{x_2} h \end{bmatrix} \begin{bmatrix} v_1 \\ -\left(\nabla_{x_2} h(x)\right)^{-1} \nabla_{x_1} h(x) v_1 \end{bmatrix} = 0 \ .$$

and therefore (12) lies in $T_x\mathcal{M}$.

In the following, we show uniqueness. Given $v_1$, assume that the vector $v$ is not unique, implying that there exists $v_2 \neq 0$ such that $\begin{bmatrix} 0 & v_2^T \end{bmatrix}^T \in \ker \nabla h(x)$. However, $\nabla_{x_2} h(x)$ is non-singular and therefore $\ker \nabla_{x_2} h(x) = \{0\}$ implying uniqueness of (12). □

**Theorem 5.** *Let $(\mathcal{M}, r)$ be a $n_1$-dimensional Riemannian manifold such that $\mathcal{M} = \{x | h(x) = 0\} \subset \mathbb{R}^{n_1+n_2}$. Consider $\mathcal{K} \subset \mathcal{M}$ and a smooth function $\phi : \mathcal{M} \to \mathbb{R}$ such that projected gradient descent is well-defined and satisfies Assumption 2.*

*Further, let $\mathcal{U} \subset \mathcal{K}$ be a non-empty (relatively) open set such that $\nabla_{x_2} h(x) = n_2$ for all $x \in \mathcal{U}$. Given the differential-algebraic system* (11)*, define for $x \in \mathcal{U}$ the control law*

$$u(x) = \pi_1[\operatorname{proj}_{\mathcal{G}}(x, -\operatorname{grad} \phi(x))] \ . \tag{13}$$

*Then, $x : [0, \epsilon) \to \mathcal{U}$ is a solution of* (11) *if and only if it is a solution of the projected gradient descent* (10)*.*

*Proof.* Let $x : [0, \epsilon) \to \mathcal{U}$ be a solution to (11) and (13). Then, by Lemma 7, $x$ has to satisfy the differential equation

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} &= \begin{bmatrix} u(x) \\ -\left(\nabla_{x_2} h(x)\right)^{-1} \nabla_{x_1} h(x) u(x) \end{bmatrix} \\ &= \operatorname{proj}_{\mathcal{K}}(x, \operatorname{grad} \phi(x)) \end{aligned}$$

and is therefore a trajectory of a projected gradient descent.

Conversely, if $x : [0, \epsilon) \to \mathcal{U}$ is a trajectory satisfying (10) then it trivially satisfies (11) with (13). □

If the set $\mathcal{K}$ is such that $\operatorname{rank} \nabla_{x_2} h(x) = n_2$ for all $x \in \mathcal{K}$ then the equivalence holds globally.

**Corollary 2.** *Let $(\mathcal{M}, r)$, $\mathcal{K}$ and $\phi$ be as in Theorem 5, i.e., such that projected gradient descent is well-defined and satisfies Assumption 2.*

*If $\mathcal{K}$ is such that $\operatorname{rank} \nabla_{x_2} h(x) = n_2$ for all $x \in \mathcal{K}$, then the control law* (13) *applied to* (11) *generates projected gradient descent trajectories that satisfy* (10)*.*

*Furthermore, if equilibrium points are isolated Corollary 1 applies and the control law* (13) *steers the system to a local minimum of $\phi$ on $\mathcal{K}$ unless the initial point is a non-minimal equilibrium.*

In conclusion, given a drift-free system that is naturally constrained to a manifold and that is regular with respect to the exogenous variables, a projected gradient descent on the feasible set can be realized by using the exogenous components of the projected gradient as a control law.

## V. Closed-Loop Power System Optimization

In this section we present an application of constrained manifold optimization in closed loop. For this, let a power network be represented by a graph $G$ with $n + 1$ nodes. Every node $i$ has an associated voltage $u_i \in \mathbb{C}$ and a power injection $s_i \in \mathbb{C}$. Node 0 acts as a voltage reference with $u_0 = 1$.

Voltages and power injections are linked algebraically by the power flow equations which are given in vector form as

$$\operatorname{diag}(u)\overline{Yu} = s, \tag{14}$$

where $Y$ is the bus admittance matrix of the grid.

### A. The Power Flow Manifold

We use the same coordinates as in [24]. That is, the state of the power system is given by $x = \begin{bmatrix} p & q & v & \theta \end{bmatrix} \in \mathbb{R}^{4n}$ where $p$ and $q$ denote the active and reactive power injection vectors (real and imaginary parts of $s$, respectively), while $v$ and $\theta$ are the vectors of voltage magnitudes and angles.

We may hence define the set

$$\mathcal{M} := \{x \in \mathbb{R}^{4n} | h(x) = 0\}$$

where $h(x)$ encodes the nonlinear equations (14), expressed in their real and imaginary parts.

It is shown in [24] that the set $\mathcal{M}$ has the structure of a $2n$-dimensional (real) smooth manifold. The tangent space of $\mathcal{M}$ at some $x \in \mathcal{M}$ is denoted as $\ker \nabla h(x)$. For an explicit expression of $\nabla h$ the reader is referred to [24].

### B. Constrained Optimization on the PFM

Next, we explicitly formulate the projected gradient descent on the power flow manifold that takes into account power generation limits and voltage constraints.

We consider the problem of minimizing power generation costs and thus define the objective function as

$$f(x) := \sum_{i=1}^{n} a_i p_i^2 + b_i p_i$$

where $a_i > 0$. Note that, while $f : \mathbb{R}^{4n} \to \mathbb{R}$ is convex in the ambient space, the notion of convexity is not a priori well-defined when $f$ is restricted to the manifold.

In addition, we consider power injection and voltage magnitude constraints at every node given by

$$\begin{aligned} \underline{p}_i &\leq p_i \leq \overline{p}_i \\ \underline{q}_i &\leq q_i \leq \overline{q}_i \\ \underline{v}_i &\leq v_i \leq \overline{v}_i \end{aligned} \tag{15}$$

for all $i = 1, \ldots, n$. In accordance with the previous sections, we express these constraints as $g(x) \leq 0$ where $g : \mathbb{R}^{4n} \to \mathbb{R}^{3n}$ and introduce the sets $\mathcal{G} := \{x \in \mathbb{R}^{4n} | g(x) \leq 0\}$ and $\mathcal{K}$ defined as in (8).

To minimize $f$ over $\mathcal{M}$ subject to the constraints (15) using projected gradient descent, we proceed as follows.

First, we endow the power flow manifold with the metric induced by the ambient space and define $\phi := f|_{\mathcal{M}}$. Consequently, $\operatorname{grad} \phi$ is given by (6). Furthermore, we assume that Assumption 1 holds for the feasible set $\mathcal{K} = \mathcal{M} \cap \mathcal{G}$. As a consequence, $\operatorname{proj}_{\mathcal{K}}$ is well-defined and the projected gradient at $x$ is the solution of the optimization problem

$$\begin{aligned} \underset{w \in \mathbb{R}^{4n}}{\text{minimize}} \quad & ||\nabla f(x) - w||^2 \\ \text{subject to} \quad & \nabla F(x) w = 0 \\ & \nabla g_{J(x)}(x) w \leq 0 \end{aligned} \tag{16}$$

where $J(x)$ the denotes the set of active constraints at $x$.

Furthermore, we assume that Assumption 2 holds for projected gradient descent on $\mathcal{K}$ and that $\mathcal{K}$ is such that $\operatorname{rank} \nabla_{x_1} h(x) = m$ and hence Corollary 2 applies.

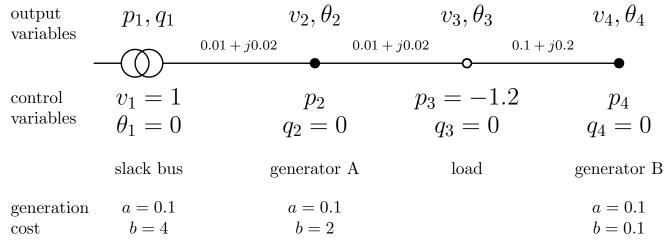


Fig. 1. 4-bus distribution grid hosting two generators and one PQ load.

Informally speaking, the operational constraints on voltage magnitude and power generation are such that the Jacobian of the power flow equations cannot become singular.

### C. Numerical Simulation

Numerical integration of a differential equation on a manifold is in general a highly non-trivial task for which sophisticated methods exist [25]. Rather than using such complex methods with superior approximation quality, we employ a basic algorithm that mimics the behavior of a real-world, discrete-time, closed-loop control system.

Let $x_0 \in \mathcal{K}$ be a feasible initial point and $\alpha > 0$ a fixed step size parameter. At each iteration performs the following two steps are performed:

(1) **Forward Euler step of exogenous variables:**

$$x_1^{n+1} = x_1^n + \alpha \pi_1 \left( \operatorname{proj}_{\mathcal{K}}(x^n, \operatorname{grad} \phi(x^n)) \right)$$

according to (16).

(2) **Natural evolution of endogenous variables:** Compute $x_2^{n+1}$ such that $x^{n+1} \in \mathcal{M}$.

Step (1) is the action performed by a discrete-time controller. Since active and reactive power injections are the only exogenous variables in our case study, this amounts to updating set-points of generators.

Step (2) is naturally performed by the physics that govern the power grid. In simulation, however, this is performed by a power flow solver that solves the power flow equations given the power injections at each node. This constitutes a natural projection on the power flow manifold $\mathcal{M}$.

Our closed-loop discrete-time control system can hence be interpreted as an instance of the numerical integration scheme based on the projection algorithm [25, Algorithm 4.2] for differential-algebraic systems of index 1.

### D. Simulation Results

In order to illustrate the application of the proposed approach to a power distribution system, we consider a simplified 4-bus grid, hosting two generators and one load (Figure 1). The substation is modeled as a slack bus, while all the remaining buses are modeled as PQ buses. An overvoltage limit of 1.05 p.u. is in force at all buses. Generator $A$ has a maximum active power limit of 0.4 MW.

The trajectory of the projected gradient flow is represented in Figure 2. The discontinuous behavior is evident: after an initial increase in the power injection of Generator B, its overvoltage limit becomes active. The trajectory remains on

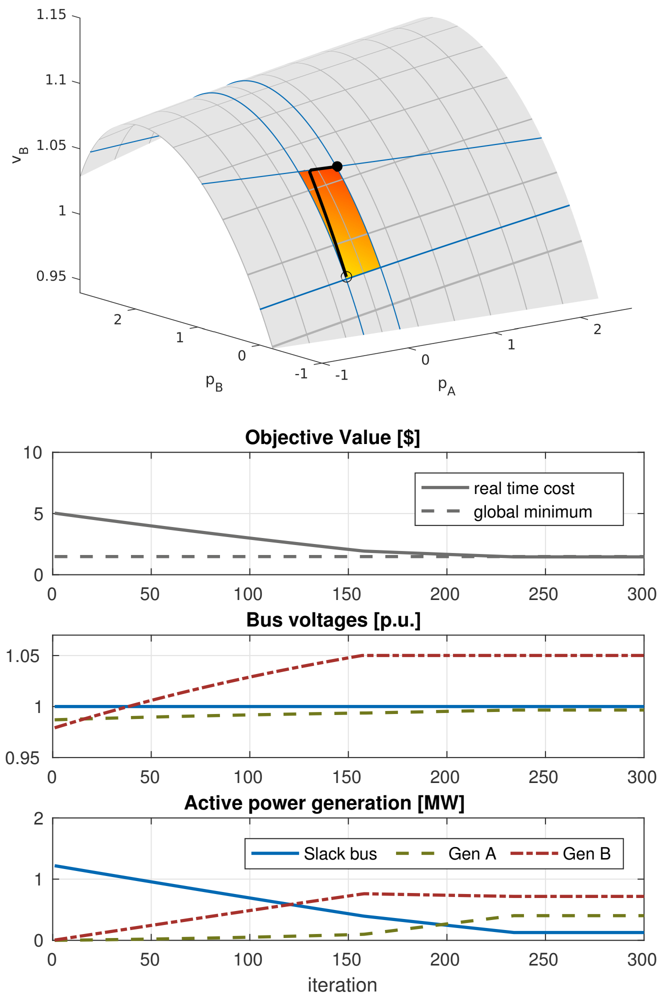


Fig. 2. **Top panel:** representation of the power flow manifold. The colored patch represents the feasible subset of the power flow manifold, defined by the operational constraints (in blue). The thick line represent the trajectory of the the system on the manifold. The empty circle and the filled dot represent the initial conditions and the attractive equilibrium, respectively. **Bottom panel:** Objective value, bus voltages, and active power generation, along the trajectory of the system.

the boundary of the feasible set until the active power limit of Generator A becomes active, at the system equilibrium (and local minimum of the cost function).

## VI. Conclusion

We have introduced a class of inequality-constrained optimization problems on manifolds and proposed a continuous-time projected gradient descent algorithm over the feasible set. Our first main result establishes convergence to equilibrium points. If the equilibria are isolated, our second result states that all minimizers are strict, and they are the only asymptotically stable equilibria.

We have shown that this approach is particularly appropriate for online load flow optimization in power systems where the state is naturally constrained to a manifold, and where operational constraints have to be enforced at all time (i.e. for the entire trajectory of the system).


## Acknowledgment

The authors would like to thank Alessandro Zanardi for helping to provide the numerical simulation results.


## References


[1] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, ser. Communications and Control Engineering. Springer London, 1996.

[2] S. T. Smith, "Optimization Techniques on Riemannian Manifolds," *arXiv:1407.5965 [cs, math]*, Jul. 2014, arXiv: 1407.5965.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[4] S. Bolognani, G. Cavraro, R. Carli, and S. Zampieri, "Distributed reactive power feedback control for voltage regulation and loss minimization," *IEEE Trans. on Automatic Control*, vol. 60, no. 4, Apr. 2015.

[5] S. Bolognani and S. Zampieri, "A distributed control strategy for reactive power compensation in smart microgrids," *IEEE Trans. on Automatic Control*, vol. 58, no. 11, Nov. 2013.

[6] N. Li, L. Chen, C. Zhao, and S. H. Low, "Connecting automatic generation control and economic dispatch from an optimization view," in *American Control Conference (ACC), 2014*. IEEE, 2014.

[7] L. Gan and S. H. Low, "An Online Gradient Algorithm for Optimal Power Flow on Radial Networks," *IEEE J. on Selected Areas in Communications*, vol. 34, no. 3, Mar. 2016.

[8] E. Dall'Anese and A. Simonetto, "Optimal Power Flow Pursuit," *arXiv:1601.07263 [math]*, Jan. 2016.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer Berlin Heidelberg, 2012.

[10] J. Dattorro, *Convex optimization & Euclidean distance geometry*. Lulu.com, 2010.

[11] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, ser. Grundlehren der mathematischen Wissenschaften. Springer Science & Business Media, 2009.

[12] J. E. Spingarn and R. T. Rockafellar, "The Generic Nature of Optimality Conditions in Nonlinear Programming," *Mathematics of Operations Research*, vol. 4, no. 4, 1979.

[13] J. M. Lee, *Introduction to Smooth Manifolds*, ser. Graduate Texts in Mathematics. Springer New York, 2012.

[14] ——, *Riemannian Manifolds - An Introduction to Curvature*, ser. Graduate Texts in Mathematics. Springer New York, 1997.

[15] A. Y. Burtscher, "Length structures on manifolds with continuous Riemannian metrics," *New York J. Math*, vol. 21, 2015.

[16] A. Bacciotti and F. Ceragioli, "Nonpathological Lyapunov functions and discontinuous Carathéodory systems," *Automatica*, vol. 42, no. 3, Mar. 2006.

[17] A. Cherukuri, E. Mallada, and J. Cortés, "Convergence of Caratheodory solutions for primal-dual dynamics in constrained concave optimization," in *Proc. of the Conf. on Control and its Applications*, Jul. 2015.

[18] A. Nagurney and D. Zhang, *Projected Dynamical Systems and Variational Inequalities with Applications*, ser. International Series in Operations Research & Management Science. Springer Science & Business Media, 1996.

[19] P. Dupuis and A. Nagurney, "Dynamical systems and variational inequalities," *Annals of Operations Research*, vol. 44, no. 1, Feb. 1993.

[20] P. Dupuis and H. Ishii, "On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications," *Stochastics and Stochastic Reports*, vol. 35, no. 1, Mar. 1991.

[21] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.

[22] P.-A. Absil and K. Kurdyka, "On the stable equilibrium points of gradient systems," *Systems & Control Letters*, vol. 55, no. 7, 2006.

[23] S. Sastry, "Geometric Nonlinear Control," in *Nonlinear Systems*, ser. Interdisciplinary Applied Mathematics. Springer New York, 1999.

[24] S. Bolognani and F. Dörfler, "Fast power system analysis via implicit linearization of the power flow manifold," in *53rd Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2015.

[25] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*. Springer Science & Business Media, 2006.