Load Variation Enables Escaping Poor Solutions of Time-Varying Optimal Power Flow

Julie Mulvaney-Kemp
University of California, Berkeley
julie_mulvaney-kemp@berkeley.edu

Salar Fattahi
University of California, Berkeley
fattahi@berkeley.edu

Javad Lavaei
University of California, Berkeley
lavaei@berkeley.edu

***Abstract*—This paper analyzes solution trajectories for optimal power flow (OPF) with time-varying load. Despite its nonconvexity, it is common to solve time-varying OPF sequentially over time using simple local-search algorithms. We aim to understand the local and global optimality behavior of these local solution trajectories. An empirical study on California data shows that local solution trajectories initialized at different points may converge to the time-varying global solution of the data-driven OPF, even if the problem has multiple local solutions throughout time. That is, these trajectories can avoid poor solutions. To explain this phenomenon, we introduce a backward mapping that relates a neighborhood of the time-varying OPF's global solution at a given time to a set of desirable initial points. We show that this proposed backward mapping could act as a stochastic gradient ascent algorithm on an implicitly convexified formulation of OPF, which justifies the escaping of poor solutions over time.**

I. Introduction

Optimal power flow (OPF) is a nationwide optimization problem that is at the core of the daily operation of power systems. OPF aims to find a cost-minimizing operating point for a power system, subject to various operational and security constraints [1]. The nonconvexity of OPF is a major impediment to its efficient and optimal solvability in practical settings. Despite this complexity, the OPF problem is solved every few minutes to match the system's power generation to a demand profile that changes over time. The inherent complexity of solving the AC model of OPF is mainly due to its nonconvex constraints which are governed by physical laws. Such nonconvexity in the problem may give rise to poor local solutions in power systems [2], [3] and in machine learning [4]. With the goal of addressing the underlying nonconvexity of the problem, a recent line of research has focused on approximating the problem as a single or sequence of convex optimization problems. These works include quadratic convex [5], second-order conic programming [6], and semidefinite programming [7], [8] relaxations.

Despite desirable theoretical guarantees, the convex relaxations of OPF suffer from two major drawbacks: 1) Their global guarantees often come at the expense of higher runtimes or overly complicated implementations; 2) They do not account for the time-varying nature of demand. This time-varying property poses additional constraints on the ramping capabilities of the generators, which in turn gives rise to a coupled optimization problem that should be solved sequentially over time. In this work, we consider the time-varying OPF with ramping constraints, where the load profile changes over time. Unlike the previous convexification techniques, we solve the problem sequentially using a simple local-search algorithm. Due to the nonconvex nature of the problem, the local-search algorithm may "become stuck" at a spurious (non-global) local solution, thus leading to a potentially large optimality gap. Previously, we made the observation on a small system in [9] that the sequence of four local solution costs could converge over time. Here, we present an extensive empirical study on a larger system with 16 spurious solutions, and show that all feasible local solution sequences (also called trajectories) converge in cost and value to the best solution. Notably, this phenomenon occurs despite the fact that the problem has multiple point-wise poor local minima at almost all times. For this system, we show that there is an *escaping period* in which different local solution trajectories converge to a solution with lowest cost, followed by a *tracking period* in which the local trajectories can closely track the global solution. In other words, *load variation enables the local solution trajectories to avoid poor solutions over time*.[1]

To explain this observation, we provide a backward-in-time mapping from a neighborhood of the globally optimal solutions of OPF at a given time (namely, end of the escaping period) to the set of desirable initial points. By leveraging its special structure, we show that the proposed backward mapping may act as a stochastic gradient ascent algorithm on an implicitly convexified formulation of the OPF problem, which in turn explains why local solution trajectories could avoid poor solutions over time.

II. Empirical Study of Time-Varying OPF

In this section, we analyze the local solution trajectories of time-varying OPF for a 39-bus system with California load data. The solution trajectories of time-varying OPF are constructed by sequentially solving a series of optimization problems with time-varying demand levels using a local-search algorithm. To prevent the solution from changing abruptly over a short period of time, the sequential optimization problems are coupled via so-called *ramping constraints*, as we explain below.

[1]Note that with constant (time-invariant) load, all the local solution trajectories will remain unchanged over time.

A. Model Details

To examine the behavior of different local solution trajectories, we consider a modified version of the IEEE 39-bus system, as introduced in [2]. Specifically, the real and reactive power demands are reduced by 50%, voltage limits tightened from +/-6% to +/-5%, and the cost functions associated with all generators are assumed to be linear. This system is known to have 16 local solutions for the fixed demand values. In this work, we take into account the time-varying nature of the load profile. In particular, the shape of the demand curve is based on the California's net load for an average day in January 2019 [10] (Fig. 1). The reported actual hourly net load data was interpolated linearly to produce a net load estimate for each 15-minute interval within 24 hours. The curve is normalized and shifted so that time 0 represents 5:00 a.m. All demands are scaled proportionally to this curve. Finally, we introduce the ramping constraints that limit the change in power generation for each generator over time. Here, the maximum magnitude of allowable change in power generation between two consecutive time steps is 5% of the capacity of each generator.

B. Behavior of Discrete Local Solutions

According to [2], the OPF for the modified IEEE 39-bus system has 16 local solutions at time $t = 0$. Starting from these initial local solutions, we constructed the sequences of local trajectories using the MATPOWER optimization toolbox and `fmincon` sequential quadratic programming solver[2] in the following procedure. We ran Algorithm 1 for all 16 initial local solutions and obtained 16 different solution sequences, which are called *discrete local trajectories* [9]. Among these trajectories, only four remained feasible throughout the span of twenty-four hours (local search may not always find a feasible point or such point may not even exist). Fig. 2 shows the point-wise distance between these four feasible trajectories and the feasible trajectory with the lowest cost (labeled as *Trajectory 1*). Interestingly, all four trajectories converge to *Trajectory 1* after seven hours.

Based on this observation, one may speculate that the problem becomes devoid of spurious local solutions over time. This is not the case for the considered problem. We uniformly searched the feasible region of the problem without ramping constraints and verified that there are multiple point-wise spurious local solutions for the static (decoupled) OPF problem at different times. In particular, there are many local solutions around the *escape time* (hour 7) when the discrete local trajectories merge into one trajectory. Fig. 3 shows the normalized objective cost values for different discrete local trajectories, alongside the costs of the discovered point-wise local solutions. Despite the existence of multiple sub-optimal operating points at different times, the discrete local trajectories initialized at various local solutions result in the lowest cost values over time. Fig. 4 takes a closer look at the active and reactive power generation for three generators. This figure shows that the problem has point-wise local solutions with a wide range of generation levels, pinpointing the importance of finding the solution with the lowest cost.

[2]Note that unlike many interior point methods that require strictly feasible initial points, `fmincon` sequential quadratic programming gives a second-order critical point even if the initial point is not strictly feasible.

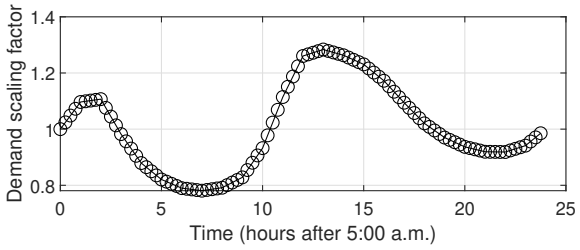

Fig. 1. Average daily net load for California during January 2019 [10]

Algorithm 1 Algorithm for obtaining discrete local trajectories

Input: Power system model with a fixed initial point $\mathbf{x}_0$
Output: Discrete local trajectory $\{\mathbf{x}_t\}_{t=0}^{K}$

1: **Initialization** : $t = 1$
2: **for** every 15-minute time increment over a span of 24 hours **do**
3: Set demand constraints for each bus based on the demand curve at time t.
4: Set generator production limits based on $\mathbf{x}_{t-1}$ and the ramping constraint.
5: Solve the corresponding OPF problem with fixed demand using `fmincon` with the initial point $\mathbf{x}_{t-1}$. Upon feasibility, collect the solution as $\mathbf{x}_t$
6: **end for**
7: **return** $\{\mathbf{x}_t\}_{t=0}^{T}$

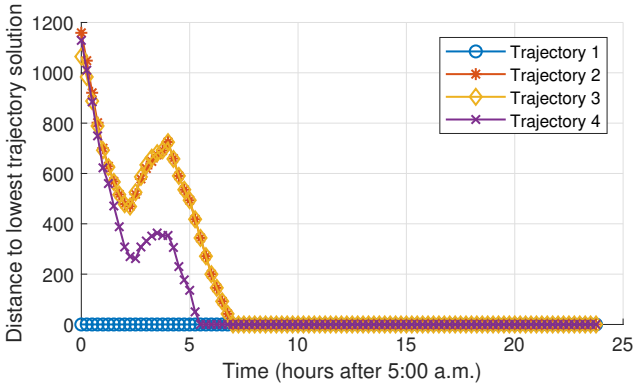

Fig. 2. Solution convergence for points on discrete local trajectories

III. Mathematical Analysis of Time-Varying OPF

The aforementioned case study reveals an important property of the time-varying OPF: In the *escaping period*, different discrete local trajectories converge to the operating point with the lowest cost. Then, in the *tracking period*, the discrete local trajectories track these globally optimal operating points, even if the load profile changes gradually over time. Such

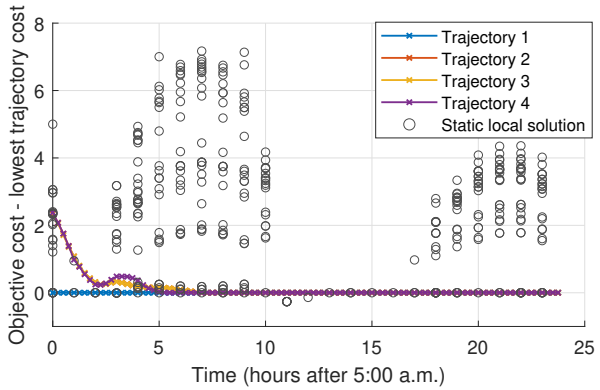

Fig. 3. Cost for points on four discrete local trajectories and point-wise local solutions (for decoupled OPF), relative to the cost of the best trajectory

tracking period has been studied in [11], [12], but the striking feature of power systems is the existence of escaping periods. To better understand this phenomenon, we will convert the time-varying OPF to an unconstrained optimization problem. Using the derived unconstrained optimization, we introduce a backward mapping that fully characterizes the dynamics of the discrete local trajectories over time. We show that the convergence of different local trajectories can be explained by the expansive property of this backward mapping. Finally, we draw a novel connection between our derived mapping and stochastic gradient ascent and use this insight to explain why the discrete local trajectories can escape poor solutions.

A. Unconstrained Model for OPF with Fixed Demand

Suppose that the considered network has n buses, each connected to a set of possibly nonzero loads and/or generators. The AC model of the OPF with fixed and predefined demand values can be written compactly as a polynomial optimization with both equality and inequality constraints [1]:

$$\min_{\mathbf{x}\in\mathbb{R}^p} f(\mathbf{x}) \tag{1a}$$

$$\text{s.t. } h_i(\mathbf{x}) = d_i, \quad i = 1, ..., n \tag{1b}$$

$$g_j(\mathbf{x}) \leq c_j, \quad j = 1, ..., m \tag{1c}$$

Here, $\mathbf{x}$ is the concatenation of voltage angles and magnitudes at different buses, as well as the active and reactive power generation outputs for different generators. ($\mathbb{R}$ is the set of real numbers.) The equality constraint (1b) ensures that the generated power meets the demand, where d_i is the demand connected to bus i, and respects the underlying structure and physical constraints of the network. The remaining constraints in the problem—including the upper and lower bounds on the voltage magnitudes and degrees, power generation, and line flows—are captured by inequality constraints (1c). We refer the reader to [1], [2] and [7] for more information on the exact formulation of the problem. Note that $f(\mathbf{x})$, $h_i(\mathbf{x})$, and $g_j(\mathbf{x})$ are polynomial functions, and hence, continuously differentiable (piecewise linear cost functions can also be reformulated as such).

It is desirable to transform OPF into an unconstrained optimization problem using the so-called implicit function theorem [13]. First, we convert (1c) to equality constraints using a set of slack variables, as in

$$\min_{\mathbf{w}\in\mathbb{R}^{p+m}} f(\mathbf{w}) \text{ s.t. } H(\mathbf{w}) = \begin{bmatrix} h_1(\mathbf{x}) - d_1 \\ \vdots \\ h_n(\mathbf{x}) - d_n \\ g_1(\mathbf{x}) - c_1 + y_1^2 \\ \vdots \\ g_m(\mathbf{x}) - c_m + y_m^2 \end{bmatrix} = 0. \tag{2}$$

where $\mathbf{w} := \begin{bmatrix}\mathbf{x}^\top & \mathbf{y}^\top\end{bmatrix}^\top$. It is easy to verify that $p > n$. Assuming that constraint qualifications hold for a given feasible point $\mathbf{w}_\star$ satisfying the Karush-Kuhn-Tucker (KKT) conditions for (2), $\mathbf{w}_\star$ can be partitioned into two sub-vectors ${\mathbf{w}_\star}^B \in \mathbb{R}^{n+m}$ and ${\mathbf{w}_\star}^R \in \mathbb{R}^{p-n}$ such that the Jacobian of $H(\mathbf{w}_\star)$ with respect to $\mathbf{w}^B$ is invertible. Therefore, the implicit function theorem guarantees the existence of a unique differentiable function ϕ such that $\mathbf{w}^B = \phi(\mathbf{w}^R)$ in a local neighborhood of $\mathbf{w}_\star$. Given such function, Problem (2) can be re-written as (see [13]):

$$\min_{\mathbf{w}^R\in\mathbb{R}^{p-n}} f(\phi(\mathbf{w}^R), \mathbf{w}^R). \tag{3}$$

Remark 1: Note that (3) *cannot be formulated explicitly, due to the unknown nature of the local solution* $\mathbf{w}_\star$ *and the function* $\phi(\mathbf{w}^R)$*. Instead, we will use this formulation as an intermediate step to analyze the behavior of the discrete local trajectories over time.*

B. Unconstrained Model for Time-Varying OPF

The above analysis reveals that, under some technical conditions, the OPF problem with fixed load is equivalent to an unconstrained optimization problem with a differentiable objective function. In this subsection, we extend our analysis to the time-varying OPF problem where demand changes over time and the problem must account for ramping constraints. As previously stated, the aim of ramping constraints is to ensure that the solution does not change too drastically from one time step to the next. One way to softly impose the ramping constraint is through a proximal method, in which the distance between the current and previous solutions is penalized in the objective function of the optimization [14]. For the time-varying OPF with K equally-spaced time steps $t_0 = 0, t_1 = \Delta t, ..., t_K = K\Delta t$ ($\Delta t > 0$), the unconstrained approximation can be written as the following sequence of optimization problems:

$$\min_{\mathbf{w}^{R_k}\in\mathbb{R}^{p-n}} f_{t_k}(\phi_{t_k}(\mathbf{w}^{R_k}), \mathbf{w}^{R_k}) + \alpha \left|\left| \mathbf{w}^{R_k} - \mathbf{w}_{\star t_{k-1}}^{R_k} \right|\right|_2^2 \tag{4}$$

for $k = 1, ..., K$, where $\alpha > 0$ is the penalization parameter, and $\mathbf{w}_{\star t_{k-1}} = \left[\left(\mathbf{w}_{\star t_{k-1}}^{B_k}\right)^\top \quad \left(\mathbf{w}_{\star t_{k-1}}^{R_k}\right)^\top\right]^\top$ is a local solution to Problem (3) at time t_{k-1}. (Note that $\mathbf{w}^{B_k}$ is not regularized without loss of generality in light of its dependence

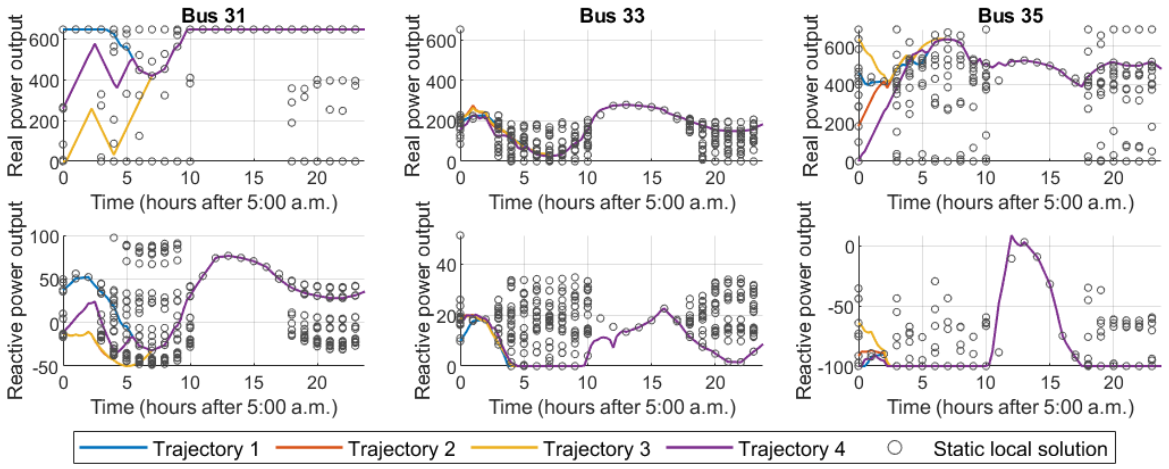

Fig. 4. Real and reactive power output of select generators: points on discrete local trajectories and point-wise local solutions

on $\mathbf{w}^{R_k}$.) Due to time-varying nature of the demand, the functions f_{t_k} and ϕ_{t_k} and the partition (R_k, B_k) change over time, and hence, they are indexed by the time steps. To simplify the analysis, assume that the partition (B_k, R_k) does not change over time, i.e., we have $B_k = B$ and $R_k = R$ for $k = 1, ..., K$. If the partition changes, then the entire time interval should be divided into sub-intervals, each with a constant partitioning of $\mathbf{w}$, and then use the argument of the next section for each sub-interval. Problem (4) can be written as

$$\min_{\mathbf{z}\in\mathbb{R}^{p-n}} \; F_k(\mathbf{z}) + \alpha \, ||\mathbf{z} - \mathbf{z}_{k-1}||_2^2 \tag{5}$$

for $k = 1, ..., K$, where $\mathbf{z} = \mathbf{w}^{R_k}$, $\mathbf{z}_{k-1} = \mathbf{w}_{\star t_{k-1}}^{R_k}$, and $F_k(\mathbf{z})$ is defined as $f_{t_k}(\phi_{t_k}(\mathbf{z}), \mathbf{z})$. This problem is strongly convex and has a unique solution if α is large.

IV. Backward Mapping

The above analysis reveals that a local-search algorithm used to solve the time-varying OPF implicitly aims to recover a stationary point of the unconstrained problem (5). Therefore, we focus on (5) in our subsequent analysis. Consider a given time $T\Delta t$, playing the role of the end of the escaping period. It is easy to see that a sequence of stationary points $\{\mathbf{z}_k\}_{k=1}^T$ for (5) satisfies the equation $\nabla F_k(\mathbf{z}_k) + 2\alpha(\mathbf{z}_k - \mathbf{z}_{k-1}) = 0$ for every $k = 1, 2, \ldots, T$ (where ∇ is the gradient operator). Therefore, given the solution $\mathbf{z}_{k-1}$, this equation defines an implicit formula for obtaining $\mathbf{z}_k$. However, going backward in time, one can write $\mathbf{z}_{k-1}$ in terms of $\mathbf{z}_k$:

$$\mathbf{z}_{k-1} = \mathbf{z}_k + \frac{1}{2\alpha}\nabla F_k(\mathbf{z}_k) := G_k(\mathbf{z}_k) \tag{6}$$

This gives rise to the following end-to-end backward mapping from $\mathbf{z}_T$ to the initial point $\mathbf{z}_0$ via the composition operator $\circ$:

$$\mathbf{z}_0 = G_1 \circ G_2 \circ \ldots G_K(\mathbf{z}_T) \tag{7}$$

Provided that the mapping (7) from $\mathbf{z}_T$ to $\mathbf{z}_0$ is expansive enough, a large set of initial points (even multiple local solutions of OPF at time 0) are guaranteed to converge to a small neighborhood of the globally optimal solution of the problem at time $t = T\Delta t$. This expansive nature of the mapping implies escaping the spurious local solutions from time 0 to $T\Delta t$. Then, the global solutions at future times will be tracked successfully if the data variation is not high [11]. This expansive property can be observed in the empirical study of the modified IEEE-39 system under California data.

A. Connection to Stochastic Gradient Ascent

This section aims to explain how data variation plays a key role in escaping spurious local solutions of time-varying OPF. Specifically, we will show that the backward mapping (6) can be treated as a variant of stochastic gradient ascent on a smoothed version of the function $F_k(\mathbf{z})$. This gives rise to the following important observation:

A certain level of stochasticity in $\{F_k(\mathbf{z})\}_{k=1}^T$ over time may enable the stationary points $\{\mathbf{z}_k\}_{k=1}^T$ to escape "sharp" local minima over time.

To explain this phenomena, first we introduce the smoothing property of the stochastic gradient descent (SGD). Recently, [15] proposed an alternative viewpoint to SGD and its ability to avoid spurious sharp local minima. Given an initial point $\mathbf{x}_0$, suppose that our goal is to find the global minimum of a (time-invariant) function $F(\mathbf{z})$ using SGD. Accordingly, the iterations of SGD can be written as

$$\mathbf{z}_{k+1} = \mathbf{z}_k - \eta(\nabla f(\mathbf{z}_k) + \omega_k), \qquad \forall k \in \{0, 1, 2, \ldots\} \tag{8}$$

where ω_t is a bounded random variable with zero mean and η is a predefined step-size. Upon defining $\tilde{\mathbf{z}}_k = \mathbf{z}_k - \eta\nabla F(\mathbf{z}_k)$, one can write the above iterations in terms of the intermediate sequence $\{\tilde{\mathbf{z}}_k\}$, as in

$$\tilde{\mathbf{z}}_{k+1} = \mathbf{z}_k - \eta\omega_k - \eta\nabla F(\tilde{\mathbf{z}}_k - \eta\omega_k), \forall k \in \{0, 1, 2, \ldots\} \tag{9}$$

To analyze the average behavior of SGD, one can consider the evolution of $\mathbb{E}_{\omega_k}(\tilde{\mathbf{z}}_{k+1})$, where the expectation is taken over ω_k conditioned on $\{\omega_0, \ldots, \omega_{k-1}\}$. Hence,

$$\mathbb{E}_{\omega_k}(\tilde{\mathbf{z}}_{k+1}) = \tilde{\mathbf{z}}_k - \eta\nabla\mathbb{E}_{\omega_k}\left(F(\tilde{\mathbf{z}}_k - \eta\omega_k)\right) \tag{10}$$

for all $t \in \{0, 1, 2, \dots\}$. Therefore, on average, SGD acts as the exact gradient descent on the surrogate function $\mathbb{E}_{\omega_k}\left(F(\tilde{\mathbf{z}}_k - \eta\omega_k)\right)$. Comparing this function with $F(\mathbf{z})$, one can verify that the former is a smoothed version of the latter, where the smoothness is due to the convolution of $F(\mathbf{z})$ with the probably density function of the random variable ω_k. Note that the original function may not be convex, and yet the convolution may become convex. As illustrated in [15], such convolution may give rise to (one-point) strong convexity of $\mathbb{E}_{\omega_k}\left(F(\tilde{\mathbf{z}}_k - \eta\omega_k)\right)$ with respect to the globally optimal solution, which in turn guarantees the convergence of $\{\tilde{\mathbf{z}}_k\}$ (and hence $\{\mathbf{z}_k\}$) to a small neighborhood around the global solution even in presence of sharp local minima. A key takeaway from this observation is that $F(\mathbf{z})$ can possess multiple sharp, poor local minima, and yet its smoothed version $\mathbb{E}_{\omega_k}\left(F(\tilde{\mathbf{z}}_k - \eta\omega_k)\right)$ may be devoid of such solutions.

Going back to the time-varying OPF and the backward mapping (6), we assume that the variation in $\{\nabla F_k(\mathbf{z})\}_{k=1}^T$ follows a stochastic process indexed by the time k. In particular, we write $\nabla F_k(\mathbf{z}) - \nabla F_{k+1}(\mathbf{z}) = \zeta_k(\mathbf{z}) + \omega_k$, where $\zeta_k(\mathbf{z})$ is a deterministic and time-varying function and ω_k is a bounded random variable with zero mean. Such assumption is realistic in power systems, where the demand response can be modeled as a deterministic time-varying function capturing its average behavior, together with an additive stochastic term accounting for its random nature. The iteration (6) is equivalent to

$$\begin{aligned}\mathbf{z}_k =& \mathbf{z}_{k+1} + \frac{1}{2\alpha}\nabla F_T(\mathbf{z}_{k+1}) \\ &+ \frac{1}{2\alpha}\sum_{\tau=k+1}^{T-1} \underbrace{\left(\nabla F_\tau(\mathbf{z}_{k+1}) - \nabla F_{\tau+1}(\mathbf{z}_{k+1})\right)}_{\zeta_\tau(\mathbf{z}_{k+1})+\omega_\tau}\end{aligned} \tag{11}$$

which can be written as the following dynamical model

$$\mathbf{z}_k = \mathbf{z}_{k+1} + \frac{1}{2\alpha}\nabla F_T(\mathbf{z}_{k+1}) + \frac{1}{2\alpha}\nu_{k+1}(\mathbf{z}_{k+1}) \tag{12a}$$

$$\nu_{k+1}(\mathbf{z}_{k+1}) = \nu_{k+2}(\mathbf{z}_{k+1}) + \zeta_{k+1}(\mathbf{z}_{k+1}) + \omega_{k+1} \tag{12b}$$

where $\nu_{k+1}(\mathbf{z}_{k+1})$ is referred to as the *variation process*. In particular, (12b) defines explicit dynamics for the variation process comprising of three parts. The first term $\nu_{k+2}(\mathbf{z}_{k+1})$ captures the correlation between the variation processes at times t_{k+1} and t_{k+2}. Furthermore, the term $\zeta_{k+1}(\mathbf{z}_{k+1})$ captures the *bias* that is added to the noise process at time t_{k+1}. Finally, ω_{k+1} is an independent randomness that is injected into the variation process at time t_{k+1}. Comparing (12) with (8), one can verify that the former reduces to stochastic gradient ascent if $\nu_{k+2}(\mathbf{z}_{k+1}) + \zeta_{k+1}(\mathbf{z}_{k+1}) = 0$. Therefore, if ω_{k+1} dominates the first two terms, (12) resembles an approximate version of stochastic gradient ascent applied to $F_T(\mathbf{z})$; otherwise, it is a biased version of GSD. Similar to (10), this implies that, on average, the points generated via the backward mapping (6) would be close to the iterations of the gradient ascent on the smoothed version of $F_T(\mathbf{z})$. This implies that, despite the possible existence of multiple spurious and sharp local minima in $\{F_k(\mathbf{z})\}_{k=1}^T$, the smoothed version of $F_T(\mathbf{z})$ may be strongly convex. This together with the expansive nature of gradient ascent on strongly convex functions [16] yields that the end-to-end backward mapping (7) is expansive, and the discrete local trajectories can escape poor local solutions over time.

V. Conclusions

This paper studies time-varying optimal power flow (OPF) problems, where a set of optimization problems should be solved sequentially due to load data variation over time. The solution to each OPF is obtained using local search initialized at the solution of the previous OPF. We offer a case study on a 39-bus system under California data, where the OPF at the initial time has 16 solutions leading to 4 feasible solution trajectories. We show that all trajectories converge to the best solution trajectory, even though OPF has many local minima at almost all times. To understand this highly desirable property, we introduce the notions of escaping period and tracking period, study the behavior of the time-varying OPF during the escaping period via a backward-in-time mapping, and relate it to a biased GSD algorithm. We show that enough data variation enables escaping poor solutions of OPF over time.

References

[1] J. A. Momoh, *Electric Power System Applications of Optimization*. Boca Raton: CRC Press, /12/19 2017. [Online]. Available: https://www.taylorfrancis.com/books/9781315218953

[2] W. A. Bukhsh, A. Grothey, K. I. M. McKinnon, and P. A. Trodden, "Local solutions of the optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4780–4788, Nov 2013.

[3] R. Y. Zhang, J. Lavaei, and R. Baldick, "Spurious local minima in power system state estimation," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 1086–1096, 2019.

[4] R. Y. Zhang, S. Sojoudi, and J. Lavaei, "Sharp restricted isometry bounds for the inexistence of spurious local minima in nonconvex matrix recovery," *Journal of Machine Learning research*, 2019.

[5] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "The QC relaxation: A theoretical and computational study on optimal power flow," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3008–3018, 2015.

[6] B. Kocuk, S. S. Dey, and X. A. Sun, "Strong socp relaxations for the optimal power flow problem," *Operations Research*, vol. 64, no. 6, pp. 1177–1196, 2016.

[7] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 92–107, 2011.

[8] S. Sojoudi and J. Lavaei, "Exactness of semidefinite relaxations for nonlinear optimization problems with underlying graph structure," *SIAM Journal on Optimization*, vol. 24, no. 4, pp. 1746–1778, 2014.

[9] S. Fattahi, C. Josz, R. Mohammadi, J. Lavaei, and S. Sojoudi, "Absence of spurious local trajectories in time-varying optimization," 2019. [Online]. Available: https://lavaei.ieor.berkeley.edu/Time_Varing_2019_1.pdf

[10] "California ISO OASIS." [Online]. Available: http://oasis.caiso.com/

[11] O. Massicot and J. Marecek, "On-line non-convex constrained optimization," *arXiv preprint arXiv:1909.07492*, 2019.

[12] Y. Tang, "Time-varying optimization and its application to power system operation," Ph.D. dissertation, California Institute of Technology, 2019.

[13] D. Bertsekas, *Nonlinear Programming*, 3rd ed. Belmont, Mass.: Athena Scientific, 2016.

[14] C. B. Do, Q. V. Le, and C. S. Foo, "Proximal regularization for online and batch learning," *ICML*, 2009.

[15] R. Kleinberg, Y. Li, and Y. Yuan, "An alternative view: When does SGD escape local minima?" *arXiv preprint arXiv:1802.06175*, 2018.

[16] E. K. Ryu and S. Boyd, "Stochastic proximal iteration: a non-asymptotic improvement upon stochastic gradient descent," *http://stanford.edu/~boyd/papers/spi.html*, 2017.