Two-Stage Load Control for Severe Under-Frequency Conditions

Lei Tang, Member, IEEE, and James McCalley, Fellow, IEEE

Abstract—A two-stage load control scheme is presented to address severe under-frequency conditions. The first stage is event-based under-frequency load shedding that guarantees fast response to the detected high consequence disturbances, for initial system protection. The control signals are obtained by solving mixed integer programming problems. After the first stage, model predictive control is triggered in the second stage for step-by-step on-line closed-loop control of interruptible loads, the amount of which is identified via solving linear programming problems. Trajectory sensitivities are used in the optimization problems in both stages to facilitate the analysis. The proposed scheme has been tested on a 9-bus system and on the New England 39-bus system.

Index Terms—Load control, under-frequency load shedding, trajectory sensitivities, interruptible loads, model predictive control, mixed integer programming.

NOMENCLATURE

A. Basic Variables and Indices

x Vector of power system state variables
y Vector of power system algebraic variables
u Vector of power system control variables
t Continuous time index
k Time series index of discretized trajectory
l Load index
n Generator index; and 0 is for the system
d Frequency limit index
[h] Subscription indicating model predictive control step

B. Time-Constant Symbols

SL Load shedding amount
CP Load importance based on cost
f_{n,st} Steady-state frequency of generator n
f_{n,base,u} Base-case steady-state frequency of generator n
f_{n,d} Frequency limit d of generator n

This work was in part supported by the Power System Engineering Research Center (PSERC) through the project “Next generation on-line dynamic security assessment”.

L. Tang and J. McCalley are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (emails: ltang@alumni.iastate.edu; jdmc@iastate.edu).

M Large positive value
B Large positive value
f_{n,ref} Reference of generator n steady-state frequency
\Delta_{n,err} Error tolerance of generator n steady-state frequency
SL_{l,max} Maximum shedding amount of load l
T_{n,d,max} Limits of generator n frequency violating limit d
H_{n} Inertia of generator n
T_{C} Control horizon of model predictive control
T_{P} Predicted time horizon of model predictive control
\delta Sampling interval of model predictive control

C. Time-Variant Symbols

x/y\_b Base-case trajectory of x/y
x/y\_u Trajectory sensitivity of x/y to u
f\_n Frequency trajectory of generator n
f_{n,base} Base-case trajectory of generator n
u_{n,d} u_{n,d}(k) = 0 if f_{n}(k) > f_{n,d}; and u_{n,d}(k) = 1 if f_{n}(k) < f_{n,d}
T_{n,d} Duration of generator n frequency violating limit d
r_{n,d} r_{n,d}(k) = u_{n,d}(k)T_{n,d}(k) to linearize the optimization
\Delta t Time-domain simulation integration time step

II. INTRODUCTION

Power systems experience under-frequency when load exceeds generation. In severe conditions, load control can contribute to balancing load and generation. Reference [1] summarized the historical severe under-frequency conditions in which load shedding is involved, including the 17,644 MW load shedding in the U.S. Northeast blackout of 2003. This paper designs a two-stage load control scheme for severe under-frequency conditions. The designed scheme consists of event-based under-frequency load shedding (UFLS) in the first stage and model predictive control (MPC) using interruptible loads in the second stage.

UFLS is a remedial action for high consequence events that involve significant frequency decline. UFLS is the last automated measure to arrest frequency decline while preventing generator trip and subsequent blackouts [2]. NERC requires that each planning coordinator develop an UFLS program implemented by transmission owners and distribution providers [3]. In a severe under-frequency condition followed by UFLS, the system frequency recovers
to a level that largely depends on the total load shedding amount rather than the load shedding locations. However, the steady-state system frequency recovery is not the only criteria for UFLS design. The system dynamics after UFLS action are different with loads interrupted in different locations. In addition, the cost of interrupting loads of the same amount but at different locations may not be the same. A well-designed UFLS program should minimize the economic loss while satisfying both the steady-state and dynamic criteria. Existing UFLS programs are considered to be “response-based” in that they respond to a measured system condition, in this case, frequency. For example, if the frequency at a location drops below 59.3 Hz for 0.3 second, there will be 10% load shedding with 0.1 second time delay [4]. UFLS programs monitor local system frequency response after the disturbances and incorporate controls to react to actual system conditions [5], [6]. The relay settings have several scales of frequency thresholds and intentional time delay, to shed certain amounts of load according to local measurements of frequency levels (f) and possibly frequency change rates (df/dt). The relay settings are designed based on operational experience and are fixed for all scenarios. Response-based strategies such as UFLS are used when gradual increase of remedial action is acceptable, but they may not be fast enough to prevent instability following severe disturbances causing large and rapid frequency excursions [7].

Modern power grids have witnessed two recent changes that motivate interest in new forms of frequency control. One change is the increased penetration of renewable energy, most of which is not contributing to automatic generation control (AGC). Reference [8] discussed the future scenarios in California with increased penetration of renewable energy and the corresponding decrease in conventional generation. Less conventional generation leads to diminished capability of providing timely generation control. In case of severe under-frequency conditions and when conventional generation is close to their capacity limits (e.g., during low wind, high load evening conditions when solar is almost zero), AGC may not have sufficient control capacity for frequency recovery, leaving the system vulnerable to subsequent disturbances.

The other change is that demand-side controllable resources have become more prevalent. Loads are already participating in the regulation market for real-time frequency regulation [9]. In [10], it was demonstrated how to include loads, especially air-conditioning loads, as spinning reserve for load curtailment during large contingencies. Similar demand-side resources suitable for frequency regulation are refrigerators, water heaters and various other household appliances that are generally available when needed. Such loads exist in the form of small and independent units that can be controlled in a reliable, robust and relatively continuous way. A controller was developed in [11] to attach to appliances for response to under-frequency in heavily loaded conditions. Reference [12] also suggested a hierarchical structure in which substation level controllers coordinate the interruptible loads under them, making possible continuous load control. In [13], a load-side primary frequency control was designed to minimize the aggregate cost of tracking an operating point.

In this paper, interruptible loads are controlled to supplement AGC. This is attractive for three reasons. First, in severe under-frequency conditions when the system needs quick frequency recovery to withstand the follow-up disturbances, load control can respond faster than generation-based AGC, which requires mechanical movement of a valve or a water gate and is designed with constrained speed to avoid following fast varying components of load [14]. Once a load control decision is made and transmitted to the local controller, the local controller can drop the loads instantaneously by opening the breaker that connects the loads. Second, load control is available at many locations throughout the grid, in contrast to generation, which is usually concentrated locationally. Third, some types of loads, particularly heating and cooling loads, have service quality that is almost insensitive to their use for control, in contrast to renewables which are often unwilling to control away from points of maximum energy extraction. For example, based on contracted agreement with customers, Southern California Edison curtailed air-conditioning loads 37 times with duration of 5-20 minutes, and received no complaint [10].

The proposed load control scheme coordinates wide-area loads for optimal control. It has two stages. The event-based UFLS (EB-UFLS) in the first stage is driven by high consequence events, such as large power plant tripping or controlled islanding. It is event-based in that it reacts immediately to pre-identified events, rather than waiting and counting the duration of the frequency below a threshold to shed the load. The UFLS actions are determined accurately through solving optimization that coordinates wide-area load resources. Event-based strategies have been widely used for remedial action design to obtain very fast action in order to maintain system integrity with minimum time delay [15-20]. Anticipated high consequence events that can cause severe under-frequency are analyzed before they occur. The corresponding remedial actions, i.e., feeder disconnection decisions, are obtained by solving mixed integer programming (MIP) problems we call load shedding
optimization (LSO). System dynamics in LSO are approximated with linear constraints based on trajectory sensitivity analysis. The designed remedial actions form a lookup table for online decision-making. In real time, the EB-UFLS program directly takes the remedial action according to the lookup table once an analyzed event is detected, to maintain system stability and protect turbine-generators.

The second stage provides an alternative to AGC considering future severe under-frequency scenarios when high penetration of renewables reduces the AGC capability to provide timely generation control. In the second stage, MPC uses optimization to design controls at each control step. MPC is closed-loop. It repetitively optimizes the control outputs to regulate system frequency with real-time measurements and an internal system dynamic model. In our design, MPC determines control signals at each control step by solving a linear programming (LP) problem we call model predictive optimization (MPO) that coordinates wide-area interruptible load resources. In this way, MPC recovers the system frequency to the rated frequency (60 Hz), coordinating with existing AGC in its control process. MPC has been widely explored in process control with slow system dynamics. MPO uses trajectory sensitivity to approximate system frequency into LP formulation efficiently solved in this paper. MPC has also been implemented in power system real-time control and proven to be an attractive control strategy, not only for load control [12], but also for AGC [21] and reactive power control [22], [23].

Compared with other optimization-based load controls for frequency recovery, such as [6] and [13], which formulate system dynamics integration into optimization constraints that consider only swing equations, LSO and MPO interpret system dynamics into linear constraints with trajectory sensitivities that are calculated in off-line analysis. Thus, the control actions can evaluate system dynamics with detailed modeling without losing online computational efficiency. The paper compares the two methods of system dynamics modeling to show the advantages of using trajectory sensitivities.

References [24], [25] introduce a frequency primary response market that coordinates various frequency controlling resources, including their pricing, as ancillary services for under-frequency conditions. The proposed two-stage load control scheme in this paper fits into the frequency primary response market design that incorporates demand response as a unique frequency controlling resource in an individual module based on dynamic simulation.

The paper is organized as follows. Section II introduces trajectory sensitivity analysis and formulates LSO for EB-UFLS. Section III introduces MPC and formulates MPO for MPC-based load control. Section IV provides the overall design of the proposed load control scheme. Section V provides the results of case studies on two test systems. Concluding remarks are provided in Section VI.

III. LSO FOR EB-UFLS

This section provides the LSO formulation for EB-UFLS using trajectory sensitivities to approximate system dynamics when parameters change.

A. Introduction to Trajectory Sensitivity Analysis

Consider power system dynamics described by differential algebraic equations (DAE) [26]

\[
\dot{x} = F(x, y, u) \\
0 = G(x, y, u)
\]

where \(x\) are state variables, such as rotor angles and speeds; \(y\) are algebraic variables, such as voltages and power; and \(u\) are control variables, such as loads. The discrete events are modeled by switching between different sets of DAE. To obtain the trajectory sensitivities of system variables \(x\) and \(y\) with respect to parameter \(u\), we take derivatives of (1) and (2) with respect to \(u\), i.e.,

\[
\dot{x}_u = F_x x_u + F_y y_u + F_u \\
0 = G_x x_u + G_y y_u + G_u
\]

Equations (3)-(4) are called augmentations of the original DAE. They generate two new sets of unknown variables:

\[
\frac{\partial \dot{x}}{\partial u} \Rightarrow x_u \quad \text{and} \quad \frac{\partial y}{\partial u} \Rightarrow y_u.
\]

The efficient calculation of the original DAE along with their augmentations is described in [26]. This paper uses a simple procedure to approximate the trajectory sensitivities as [27]

\[
x_u = \lim_{\Delta u \to 0} \frac{x(u + \Delta u) - x(u)}{\Delta u} \approx \frac{x(u + \Delta u) - x(u)}{\Delta u}
\]

which involves simulation of the system model for parameter \(u\) and \(u + \Delta u\) where \(\Delta u\) is infinitesimal increment of \(u\). \(y_u\) can be obtained using the same method. Trajectory sensitivity linearizes the system along a nominal trajectory, rather than around an equilibrium point. Reference [28] summarized the wide applications of trajectory sensitivity analysis in the power system area. One useful function of trajectory sensitivity is to approximate the trajectory deviations after a slight change of parameters. Given base-case time-domain simulation trajectories \(x'(t), y'(t)\) and parameter change \(\Delta u\), using Taylor series expansion, the system variable...
trajectories are approximated by ignoring higher order terms, according to
\[ x(t) = x^h(t) + x(t) \quad y(t) = y^h(t) + y(t) \]
\[ \Delta x(t) = \Delta x(t) \quad \Delta y(t) = \Delta y(t) \]
Equations (7)-(8) are widely used in power system dynamic security assessment to improve computational efficiency [29-32]. This paper uses (7)-(8) to design the two-stage load control scheme.

**B. LSO Formulation**

LSO determines optimal load shedding solutions for each contingency to satisfy system performance criteria. Performance requirements for frequency dynamics are imposed within the LSO constraints. The formulation of LSO that determines the optimal load shedding (SL) amount is as follows:

\[ \text{minimize } (CP_1 \times SL_1) \quad \text{subject to:} \]

- **Frequency dynamics approximation:**
  \[ f_n(k) = f_{n,b} + \sum_{i=1}^{L} \partial f_n u \cdot SL \]
  \[ f_{n,t} = f_{n,b} + \sum_{i=1}^{N} \partial f_n t \cdot SL \]

- **Under-frequency time limits:**
  \[ U_{n,d}(k) = \left( f_n(k) - f_{n,b} \right) / M \]
  \[ U_{n,d}(k) = \left( u_{n,d}(k) - U_{n,d}(k) + 1 \right) \]
  \[ 0 \quad n, d \quad (k 1) \quad B \]
  \[ 0 \quad T_{n,d}(k 1) \quad r_{n,d}(k 1) \quad B \quad u_{n,d}(k 1) \]
  \[ T_{n,d}(k) = r_{n,d}(k - 1) + u_{n,d}(k - 1) \cdot \Delta t(k - 1) \]
  \[ T_{n,d}(k) \leq T_{n,d,max} \]

- **Reference frequency constraints:**
  \[ f_{n,ref} - \Delta f_{n,ext} \leq f_{n,t} \leq f_{n,ref} + \Delta f_{n,ext} \]

- **Load shedding constraints:**
  \[ 0 \leq SL_i \leq SL_{l,max} \]

where \( k \in \{1, \ldots, K_s\} \) is the time series index of discretized trajectory \( n \); \( n \in \{0, \ldots, N\} \) is the generator index where 0 is used to indicate the system frequency; and \( d \in \{0, \ldots, D\} \) is the time limit threshold index.

The decision variable \( SL_i \) in the objective function (9) is load shedding amount on load \( l \). The coefficients \( CP_i \) indicate the load importance based on cost, which is determined according to the demand bids in the day-ahead market [33].

UFLS involves dropping blocks of loads by disconnecting feeders. This discreteness creates modeling difficulties since feeder flows cannot be known exactly until real-time. Modeling load shedding amount as continuous variables \( SL_i \) in LSO requires on-line decisions of selecting feeders to disconnect such that the total load shedding on a bus is closest to but more than the calculated optimal values that meet the minimum performance requirements. This over-shedding is considered by MPC-based load control in stage 2, using interruptible loads.

In (10)-(11), which approximate frequency dynamics, the frequency response of generator \( n \) is approximated using trajectory sensitivities according to (7)-(8). Variable \( f_{n,b} \) is the base-case frequency along which the trajectory sensitivities are calculated. The system frequency \( f_0 \), which is used to evaluate the system response, is the center-of-inertia frequency, expressed in (20)

\[ f_0(k) = \frac{\sum_{n=1}^{N} H_n \cdot f_n(k)}{\sum_{n=1}^{N} H_n} \]

where \( H_n \) is the inertia of generator \( n \). The frequencies experience multiple swings of oscillation with decreased magnitudes before reaching their steady states. Constraints (10)-(17) only consider the early swings with the most severe oscillations. This reduces LSO’s computational burden through reducing \( K_n \) which is the total number of time steps, and thus reduces the number of constraints in LSO. Then the steady-state frequencies (11) are constrained in (18).

The frequency dynamics are constrained in (10)-(17). The post-contingency frequency response cannot exceed the threshold \( f_{n,t} \) for more than a certain time duration \( \Delta t_{n,t,\text{max}} \). Different frequency thresholds have different time limits, as indicated in Table 1. Table 1 lists two sets of time limits in different columns. One set is “time limits for generators” that avoid undesirable generator trip, as considered in [2]. The other set is “time limits for system” that guarantee acceptable performance of system frequency \( f_0 \). The system performance criteria used by Northeast Power Coordinating Council (NPCC) are used in our case studies [34].

<table>
<thead>
<tr>
<th>Index (d)</th>
<th>Frequency Limits (Hz)</th>
<th>Time Limits for Generator</th>
<th>Time Limits for System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Below 59.5</td>
<td>3 min</td>
<td>Safe</td>
</tr>
<tr>
<td>2</td>
<td>Below 58.5</td>
<td>30s</td>
<td>10s</td>
</tr>
<tr>
<td>3</td>
<td>Below 58.0</td>
<td>7.5s</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Below 57.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To evaluate the duration, a set of binary variables $u_{n,d}(k)$ indicates if the frequency is below the threshold $f_{n,d}$ or not, as in (13). The variable $f_{n,d}$ is frequency threshold $d$ of generator $n$; and $M$ is a large positive value (we have used $M = 100$) such that

$$\left| \frac{f_{n,d} - f_n(k)}{M} \right| \leq 1.$$  \hspace{1cm} (21)

Thus, $u_{n,d}(k) = 0$ if $f_n(k) > f_{n,d}$ and $u_{n,d}(k) = 1$ if $f_n(k) < f_{n,d}$. Then $T_{n,d}(k)$, duration below $f_{n,d}$ is computed as

$$T_{n,d}(k) = u_{n,d}(k-1) \cdot T_{n,d}(k-1) + u_{n,d}(k-1) \cdot \Delta t(k-1)$$  \hspace{1cm} (22)

where $\Delta t$ is integration time step. Once the frequency is higher than $f_{n,d}$, $T_{n,d}(k)$ is reset to zero. The time-domain simulation uses an implicit trapezoidal integration method [35] with variable time steps. At each time step of the integration, the step size is automatically determined considering both the simulation accuracy and the computational demand [36].

In order to avoid the nonlinearities in (22), the term $u_{n,d}(k-1)T_{n,d}(k-1)$, is re-defined as $r_{n,d}(k-1)$ to satisfy (14)-(15) where $B$ is chosen large enough (we have used $B = 100$) such that

$$B \geq T_{n,d,\max}.$$  \hspace{1cm} (23)

where $T_{n,d,\max}$, which can be found in Table I, is the time limit of threshold $d$ for generator $n$. This strategy was also deployed in [6] for setting UFLS relays.

In this way, $r_{n,d}(k-1)$ is equivalent to $u_{n,d}(k-1)T_{n,d}(k-1)$ because

$$r_{n,d}(k-1) = \begin{cases} T_{n,d}(k-1) & u_{n,d}(k-1) = 1 \\ 0 & u_{n,d}(k-1) = 0 \end{cases}$$  \hspace{1cm} (24)

After the linearization, (22) becomes (16).

The steady-state frequencies should reach the reference frequency, within a tolerance, as in (18) where $f_{n,ref}$ is the reference, and $f_{n,ref}$ is the tolerance.

IV. MPC-BASED LOAD CONTROL

A. Introduction to MPC

In MPC, a sequence of controls optimizes the predicted behavior of the monitored system variables [37]. This closed-loop control process is shown in Fig. 1. The central feature of MPC is the incorporation of an explicit process model into the control optimization, which enables the controllers, in principle, to predict and regulate the system process dynamics, such as frequencies in our case.

MPC is conventionally implemented using linear system models. In our case, the system model is nonlinear, so we convert the system frequency behaviors into linear constraints in MPO using trajectory sensitivities. MPC based on trajectory sensitivity has been implemented in the power systems area before [38].

An introduction to MPC for nonlinear systems and the standard optimization formulation can be found in [39], [40]. In the MPC optimization, $T_p$ is the prediction horizon, and $T_c$ is the control horizon. At time $t_k$, MPC solves the optimal control problem over a finite predicted time horizon $[t_k, t_k+T_p]$, to design control output $u$ for $[t_k, t_k+T_C]$ (where $T_c \leq T_p$), according to the following:

$$\text{minimize } \int_{t_k}^{t_k+T_p} F(x(\tau), y(\tau), u(\tau)) d\tau$$  \hspace{1cm} (25)

subject to:

$$\begin{align*}
\dot{x} &= f(x(\tau), y(\tau), u(\tau)), \quad x(t_k) = \tilde{x}(t_k) \quad (26) \\
0 &= g(x(\tau), y(\tau), u(\tau)), \quad y(t_k) = \tilde{y}(t_k) \quad (27) \\
n(\tau) &\in U, \quad \forall \tau \in [t_k+t_c + T_p] \quad (28) \\
u(\tau) &= u(t_k, t_k+T_c), \quad \forall \tau \in [t_k+t_c, t_k+T_p] \quad (29) \\
x(\tau) &\in X, \quad y(\tau) \in Y, \quad \forall \tau \in [t_k, t_k+T_p] \quad (30)
\end{align*}$$

The function $F(\cdot)$ evaluates the control performance which in our case is the cost of load control. The tildes ($\tilde{\cdot}$) in the formulation indicate real-time measurements as opposed to internal model variables, where the difference between the two is a result of the inaccuracies in the model and/or the model parameters. The optimization is solved for every sampling interval $\delta (\delta \leq T_c)$ when real-time measurements are obtained and used in the formulation. The optimized control $u$ are applied only for $[t_k, t_k+\delta]$, since updated control will be obtained from solving the next optimization problem at time $t_{k+1}=t_k+\delta$, based on real-time measurements at that time, with $t_k+T_c$ and $t_k+T_p$ shifted forward to $t_{k+1}+T_C$ and $t_{k+1}+T_p$.

The on-line numerical solution of the nonlinear optimization of (25)-(30) can be computationally expensive for large systems, since DAEs are involved. Approximation
using trajectory sensitivities reduces the computational burden by replacing (26)-(27) with (7)-(8). This approximation has also been deployed in MPC-based power system load control [41-45] and reactive power compensation for voltage stability [23], [44].

B. MPO Formulation for MPC-based Load Control

After the frequencies have recovered from the large excursions in the first stage, MPC in the second stage neglects binary variables that count durations in LSO and formulates MPO into LP problems for online computing. Since the second stage aims to recover the system frequency level, MPO regulates only steady-state system frequency based on trajectory sensitivity analysis, neglecting the dynamics of generator frequencies.

The MPO formulation for MPC is

\[
\begin{align*}
\min_{t=1}^L & \left( CP_l^{[h]} \times SL_l^{[h]} \right) \\
\text{subject to:} & \\
\quad & f_{0,ref}^{[h]} = f_{0,base,ut}^{[h]} + \sum_{i=1}^L \frac{\partial f_{0,i}^{[h]}}{\partial SL^{[h]}} \cdot SL_i^{[h]} \\
\quad & SL_{i,min}^{[h]} \leq SL_i^{[h]} \leq SL_{i,max}^{[h]} 
\end{align*}
\]

where superscript \([h]\) is the index of control steps. The cost-oriented coefficients \(CP_l^{[h]}\) are determined by real-time pricing, or incentives paid to participating customers [46]. For example, Midcontinent Independent System Operator (MISO) introduces in [47] how to control the demand resources in its demand response program. The implementation details including pricing can be found in the MISO demand response Business Practices Manual [48]. In our case we used identical coefficients. The reference frequency \(f_{0,ref}^{[h]}\) does not stay the same throughout the control process. We use a changing reference strategy whereby the reference frequency \(f_{0,ref}^{[h]}\) starts from the steady-state frequency before the initialization of MPC and increases by \(\Delta f_{0,inc}^{[h]}\) at each MPC step until the reference frequency reaches 60 Hz. This changing-reference strategy, which is also employed in [38] for controlling shunt capacitors for voltage recovery, is implemented to avoid large load interruption at each control step. The base-case system steady-state frequency \(f_{0,base,ut}^{[h]}\) is real-time system frequency at the time of MPO formulation. Setting \(SL_{i,min}^{[h]} < 0\) allows MPO to recover interruptible loads in over-frequency conditions. The sensitivity of system steady-state system frequency to load \(\frac{\partial f_{0,i}^{[h]}}{\partial SL^{[h]}}\) is not associated with any particular load \(l\), because the load changes of the same amount but at different locations equally contribute to the steady-state system frequency recovery.

MPO (31)-(33) can be solved heuristically using an iterative method. The total amount of load control is determined by reformulating (32) into

\[
\sum_{i=1}^L SL_i^{[h]} = f_{0,ref}^{[h]} - f_{0,base,ut}^{[h]} + \frac{\partial f_{0,i}^{[h]}}{\partial SL_i^{[h]}}.
\]

Since the interruptible loads of the same amount but at different locations equally contribute to the system steady-state frequency recovery, at each iteration MPO chooses the most economical action, i.e., fully curtailing the available load \(l\) with the smallest coefficient \(CP_l^{[h]}\) while satisfying load \(l\) limit (33). The iteration ends with partial curtailment to satisfy (34).

Fig. 2 illustrates the intended post-contingency frequency response under MPC. The controls are determined by solving (31)-(33). After the application of the designed control at \(t_k\), the real-time measurements differ from the predicted system response due to inaccuracies in the model and/or the model parameters. In the next control step at \(t_{k+1}\), real-time measurements are used to initialize (32), and this process is repeated until the control objective is achieved. The parameter \(\delta\) is chosen such that the frequency reaches a new steady state before occurrence of the next load curtailment. Using the changing-reference strategy, each MPO predicts system frequency for the next sampling interval \(\delta\), and the control is also designed for the next sampling interval \(\delta\), thus, \(T_p = T_C = \delta\) holds in our case for frequency regulation.

![MPC-based load control for frequency regulation](image)

Fig. 2 MPC-based load control for frequency regulation.

V. IMPLEMENTATION OF TWO-STAGE LOAD CONTROL

Fig. 3 shows the time frames of all controls for frequency regulation.
The dotted lines indicate the effect of the scheme on original primary (governor) and secondary (AGC) control. In real-time operation, the designed EB-UFLS is applied immediately following the contingency detection. Governor response begins at the same time. MPC begins when the system reaches the steady state after EB-UFLS and governor response. From Fig. 3, EB-UFLS reduces the amount of governor action required, and MPC reduces the amount of AGC action required. Since both MPC and AGC are closed-loop control, they can coordinate with each other to reach the reference frequency. Several factors determine the duration of MPC, including frequency reference and length of sampling interval $\delta$ [10], [12].

The flowchart to implement the proposed two-stage load control scheme is shown in Fig. 4. Off-line analysis solves LSO for identified contingencies under forecasted operating conditions if any time limit in Table I is violated. The designed controls form a lookup table that is stored in the dynamic database for online decision-making. MPC is implemented to correct the power imbalance existing after the governor control. The trajectory sensitivities for MPO-based MPC are calculated off-lines to improve the computational efficiency of online MPC. The two stages coordinate with two different time-scales for frequency recovery in severe under-frequency conditions. EB-UFLS guarantees fast response with open-loop control. Due to the discreteness of feeder disconnection and open-loop nature in EB-UFLS, EB-UFLS under-sheds and is followed by MPC. MPC provides refined closed-loop control and coordinates with existing AGC.

VI. CASE STUDIES

The goal of the case studies is to illustrate the performance of the proposed load control scheme in terms of frequency recovery, governor response and AGC response. Two systems are used to test the proposed scheme: A 9-bus system and the New England 39-bus system. A MATLAB-based program PSAT [49] performs time-domain simulation for the test systems. The mathematical expression of the simulation models can be found in [50]. The machines use a 4th order model; the turbine governors use Type II model; the prime movers are constants; and the loads use ZIP (polynomial) models with frequency sensitivities. Since the time-domain simulation and trajectory sensitivity calculation are needed only in off-line analysis, the complexity of dynamic models does not affect the online computing in real-time operation.

A. 9-Bus System

The one-line diagram of a 9-bus system, which has 6 generators, is shown in Fig. 5. Two contingencies are analyzed; and each is to trip one generator, as described in Table II.

![Fig. 3 Frequency control time frames.](image)

![Fig. 4 Flowchart of the two-stage load control scheme implementation.](image)
After tripping one generator at 0.2 second, due to the imbalance between the load and the generation, the system experiences sudden decline of generator speeds. The governors react to the contingency based on their droop characteristic [51], such that after the contingency, the frequencies recover to a level lower than the rated 60 Hz. Fig. 6 shows the system frequencies following the two contingencies. For contingency 1, all frequencies meet the performance requirements in Table I, thus only MPC-based load control is needed. For contingency 2, the steady-state frequencies are below 59.5 Hz, which does not meet the performance requirements in Table I, thus both stages are needed in load control.

EB-UFLS is designed for contingency 2. MPC is triggered for both contingencies 1 and 2 after 15 seconds, at the end of the transient period. The MPC control the interruptible loads step by step to recovery the system steady-state frequency. Using the changing reference strategy, the reference frequency \( f_{n, \text{ref}}^{(b)} \) starts from the steady-state frequency before the initialization of MPC and increases by \( \Delta f_{n, \text{inc}}^{(b)} = 0.1 \) Hz at each MPC step. When the frequencies exceed 59.9 Hz, increasing \( f_{n, \text{ref}}^{(b)} \) by 0.1 Hz will cause over adjustment, thus \( f_{n, \text{inc}}^{(b)} \) is reduced such that \( f_{n, \text{ref}}^{(b)} \) is 60 Hz. MPO recovers interruptible loads in case of over-frequency due to over-adjustment or operating condition change. The sampling interval \( \delta \) (which is 20 seconds in this case) is chosen to be large enough for transients to cease after each control action. MPC remains in effect until the system frequency is recovered to 60 Hz.

AGC signals are applied every 1s. The load control scheme is tested under two conditions. The first condition has sufficient reserve for AGC-based frequency regulation. The second condition assumes high penetration of renewable energy, when conventional generation is close to their capacity due to lack of renewable generation, as discussed in section I, thus AGC reaches the limit before the frequency recovery. The AGC ramp rates are determined in accordance with NERC operating guide [52], which requires that the frequency is back to the rated value within 10 minutes in contingency conditions. Historical severe events have shown that frequency recovery could take more than 10 minutes, e.g., the WECC event on June 14, 2004 took 18 minutes to recover [53].

In the simulation, to reflect the discreteness of load blocks in reality, as discussed in section III-B, the EB-UFLS over-sheds by a random percentage within [0%, 5%] at each bus. To account for the time needed to detect, communicate, process, and decide, the action of EB-UFLS has 0.2-second delay after the contingency. This delay is captured in the off-line analysis by the time-domain simulation and trajectory sensitivity calculation as a switching event in (1) -4. The simulation stops when the system steady-state frequency is within the range of [59.99, 60.01] Hz. There is no over-frequency condition that needs interruptible load recovery in this case. For contingency 1, the load curtailments in MPC, with sufficient AGC and limited AGC, are shown in Table III. For contingency 2, the EB-UFLS (including the over-shedding) and load curtailments in MPC, with sufficient AGC and limited AGC, are shown in Table IV. The curtailment percentages of the total system load are also listed in both tables.

**TABLE II**

<table>
<thead>
<tr>
<th>Contingency #</th>
<th>Description</th>
<th>Gen.</th>
<th>% of System Gen.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contingency 1</td>
<td>Trip generator 5</td>
<td>16 MW</td>
<td>2.78%</td>
</tr>
<tr>
<td>Contingency 2</td>
<td>Trip generator 6</td>
<td>22 MW</td>
<td>3.83%</td>
</tr>
</tbody>
</table>

**Fig. 6** Frequency responses after tripping generators. Plot (a) corresponds to generator frequencies after the two contingencies. Plot (b) corresponds to system frequencies in contingency 2 with and without EB-UFLS.
The system frequencies for the two contingencies are shown in Fig. 7 and Fig. 8. For comparison, Fig. 7 and Fig. 8 also show the effect of AGC without MPC. It is observed that, MPC recovers the system frequency to 60 Hz much faster than the situation with only AGC. In case of insufficient reserve, as in Fig. 7(b) and Fig. 8(b), AGC loses the capability of frequency regulation when it reaches the limit.

To show the advantage of using trajectory sensitivities in the proposed load control scheme, we compare the dynamics constraints (10)-(11) with the discretized swing equations that neglect controls such as voltage regulator and only consider loads of constant power. In the comparison, the discretized swing equations, as introduced in [6], replace constraints (10)-(11) in LSO. The re-formulated optimization (which we call opt-UFLS) with discretized swing equations determines control action that results in unacceptable frequency response, as shown in Fig. 9. Though opt-UFLS considers the time limits in Table I, the system dynamic behavior expressed with discretized swing equations is different from the simulation result obtained from the method based on trajectory sensitivity analysis, which is more flexible with system modeling complexity.

### Table III

<table>
<thead>
<tr>
<th>Steps</th>
<th>MPC-1</th>
<th>MPC-2</th>
<th>MPC-3</th>
<th>MPC-4</th>
<th>MPC-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>15</td>
<td>35</td>
<td>55</td>
<td>75</td>
<td>95</td>
</tr>
<tr>
<td>Curt. MW*</td>
<td>2.770</td>
<td>2.752</td>
<td>2.741</td>
<td>2.740</td>
<td>2.251</td>
</tr>
<tr>
<td>% of Load*</td>
<td>0.506%</td>
<td>0.504%</td>
<td>0.502%</td>
<td>0.502%</td>
<td>0.412%</td>
</tr>
<tr>
<td>Curt. MW**</td>
<td>2.769</td>
<td>2.749</td>
<td>2.743</td>
<td>2.752</td>
<td>2.319</td>
</tr>
<tr>
<td>% of Load**</td>
<td>0.505%</td>
<td>0.503%</td>
<td>0.502%</td>
<td>0.504%</td>
<td>0.425%</td>
</tr>
</tbody>
</table>

* Sufficient AGC; ** limited AGC

### Table IV

<table>
<thead>
<tr>
<th>Steps</th>
<th>EB-UFLS</th>
<th>MPC-1</th>
<th>MPC-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>0.4</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>Curt. MW*</td>
<td>2.047</td>
<td>3.795</td>
<td>3.775</td>
</tr>
<tr>
<td>% of Load*</td>
<td>0.374%</td>
<td>0.695%</td>
<td>0.691%</td>
</tr>
<tr>
<td>Curt. MW**</td>
<td>2.046</td>
<td>3.801</td>
<td>3.783</td>
</tr>
<tr>
<td>% of Load**</td>
<td>0.374%</td>
<td>0.696%</td>
<td>0.693%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Steps</th>
<th>MPC-3</th>
<th>MPC-4</th>
<th>MPC-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>55</td>
<td>75</td>
<td>95</td>
</tr>
<tr>
<td>Curt. MW*</td>
<td>3.725</td>
<td>3.725</td>
<td>1.351</td>
</tr>
<tr>
<td>% of Load*</td>
<td>0.682%</td>
<td>0.682%</td>
<td>0.248%</td>
</tr>
<tr>
<td>Curt. MW**</td>
<td>3.721</td>
<td>3.721</td>
<td>1.401</td>
</tr>
<tr>
<td>% of Load**</td>
<td>0.681%</td>
<td>0.681%</td>
<td>0.256%</td>
</tr>
</tbody>
</table>

* Sufficient AGC; ** limited AGC

**B. New England 39-Bus System**

The one-line diagram of the New England 39-bus system is shown in Fig. 10. The simulated contingency is to trip the
generator on bus 31 at 0.2 second. Before the contingency, the tripped generator was operating at 368.9 MW, 5.10% of the total generation. The frequencies of all generators without the proposed load control are shown in Fig. 11(a), and the corresponding system frequency is shown in Fig. 11(b).

From Fig. 11, without load control the system frequency cannot reach 59.5 Hz in 30 seconds, as required in Table I. To mitigate the impact after the contingency, off-line analysis solves LSO to determine the optimal amount of load to shed. Fig. 11(b) compares the system frequencies with and without EB-UFLS.

After the application of the EB-UFLS, the system frequency comes to a secure level as shown in Fig. 11(b). MPC is triggered after 30s when the frequencies almost reach their steady states. MPC curtails the load step by step to recover the system steady-state frequency. It is found that a 20-second sampling interval also works well for this case. As in the 9-bus case, the reference frequency $f_{n,ref}^{[h]}$ increases by $\Delta f_{n,inc}^{[h]} = 0.1$Hz at each MPC step. When the frequencies reach over 59.9 Hz, $\Delta f_{n,inc}^{[h]}$ is reduced such as $f_{n,ref}^{[h]}$ is 60 Hz. The simulation stops when the system steady-state frequency is within the range of [59.99, 60.01] Hz. There is no over-frequency condition that needs interruptible load recovery in this case. The EB-UFLS (including the over-shedding) and load curtailments in MPC, with sufficient AGC and limited AGC, are shown in Table V. The load curtailment percentages of the total system load are also listed in the table. Fig. 12 shows the system frequency simulation results.

![Fig. 10 One-line diagram of the New England 39-bus system.](image)

![Fig. 11 Frequency response comparison. Plot (a) corresponds to generator frequencies without load control. Plot (b) corresponds to comparison of system frequency with and without EB-UFLS.](image)

**TABLE V**

<table>
<thead>
<tr>
<th>Steps</th>
<th>EB-UFLS</th>
<th>MPC-1</th>
<th>MPC-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>0.4</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>Curt. MW*</td>
<td>99.41</td>
<td>48.29</td>
<td>48.21</td>
</tr>
<tr>
<td>% of Load*</td>
<td>1.417%</td>
<td>0.688%</td>
<td>0.687%</td>
</tr>
<tr>
<td>Curt. MW**</td>
<td>99.36</td>
<td>48.32</td>
<td>48.19</td>
</tr>
<tr>
<td>% of Load**</td>
<td>1.416%</td>
<td>0.689%</td>
<td>0.687%</td>
</tr>
</tbody>
</table>

* Sufficient AGC; ** limited AGC

![Fig. 12 System frequencies.](image)
From the case studies on the two test systems, the load control scheme meets the control performance as required in Table I, regardless of the limit on AGC. MPC curtails the load gradually in closed loop and reacts faster than AGC.

The online computing time for MPO is important. Unlike LSO that requires integer variables to model the system dynamic details, MPO is formulated as an LP optimization regulating system steady-state frequency recovery. The number of variables in MPO is the number of interruptible loads; and the constraints include an equation to approximate the system steady-state frequency and the limits on load control amount. This LP problem is solved heuristically using an iterative method. After all the interruptible loads are sorted only once based on $CP^{(t)}$ in (31), each iteration has the same amount of calculation regardless of system size and complexity; and the sorting can be done before MPO with known $CP^{(t)}$. The MPO computing time is proportional to the number of interruptible loads that are changed in the MPO solution. In the worse situation, the MPO computing time is proportional to the number of the total interruptible loads. In MATLAB for Windows environment with IntelCore 2 Duo CPU T6500 and 4GB RAM, the average computing time of MPO is 0.0008 second for each iteration. It is suggested to have MPO computing time within 10% of the sampling interval $\delta$, which is 20 seconds in our case studies for both systems. The MPO computing time for the two test systems is much less than the suggested 2-second limit. MPO can meet the 2-second limit with a maximum of 2500 interruptible loads in the worst situation.

If the MPO computing time is limited by existing computing resources for very large systems with more than 2500 interruptible loads, the sampling interval $\delta$ can be enlarged. In addition, a large system can be scaled into multiple areas to implement decentralized MPC that further improves the computational efficiency [54]. Reference [55] implements a decentralized MPC for load frequency control in an interconnected power system. Each area in the interconnected power system has its own local model predictive controller, which is designed independently and can communicate with other areas. The decentralized MPC has also been implemented in AGC to improve computation efficiency for large systems [56].

VII. CONCLUSIONS

This paper has presented a two-stage load control scheme. The load control solutions are obtained by solving LSO and MPO, respectively, coordinating wide-area load resources. Interruptible loads participate in the second stage of the load control scheme. Trajectory sensitivities interpolate the frequency trajectory variations with the change of loads for better computational efficiency. The EB-UFLS guarantees fast response to severe under-frequency conditions to maintain system stability with minimum time delay. With the weakened AGC-based frequency regulation due to high penetration of renewables, MPC using interruptible loads in the second stage can supplement AGC for quick frequency recovery.

VIII. REFERENCES
