Voltage Instability: Phenomena, Countermeasures, and Analysis Methods

THIERRY VAN CUTSEM, MEMBER, IEEE

Invited Paper

A power system may be subject to (rotor) angle, frequency or voltage instability. Voltage instability takes on the form of a dramatic drop of transmission system voltages, which may lead to system disruption. During the past two decades it has become a major threat for the operation of many systems and, in the prevailing open access environment, it is a factor leading to limit power transfers. The objective of this paper is to describe voltage instability phenomena, to enumerate preventive and curative countermeasures, and to present in a unified and coherent way various computer analysis methods used or proposed.

Keywords—Bifurcations, load dynamics, nonlinear systems, power systems, security analysis, stability.

I. INTRODUCTION

The transfer of power through a transmission network is accompanied by voltage drops between the generation and consumption points. In normal operating conditions, these drops are in the order of a few percents of the nominal voltage. One of the tasks of power system planners and operators is to check that under heavy stress conditions and/or following credible events, all bus voltages remain within acceptable bounds.

In some circumstances, however, in the seconds or minutes following a disturbance, voltages may experience large, progressive falls, which are so pronounced that the system integrity is endangered and power cannot be delivered correctly to customers. This catastrophe is referred to as voltage instability and its calamitous result as voltage collapse. This instability stems from the attempt of load dynamics to restore power consumption beyond the amount that can be provided by the combined transmission and generation system.

In an increasing number of systems, voltage instability is recognized as a major threat for system operation, at least as important as thermal overload and angle instability problems, known for a longer time. Several factors have contributed to this situation. It is well known that the building of new transmission and generation facilities is more and more difficult, often delayed and sometimes impossible. The building of larger, remote power plants has decreased the number of voltage controlled points and increased the electrical distance between generation and load (although this might be partially compensated by the emergence of independent power productions closer to loads). The heavy use of shunt compensation to support the voltage profile allows larger power transfers but brings the instability point closer to normal values. Also, voltage instability is often triggered by the tripping of transmission or generation equipments, whose probability of occurrence is relatively large (compared for instance to the three-phase short-circuit considered in angle stability studies). Last but not least, the transmission open access environment has created an economical incentive to operate power systems closer to their limits. More than ever, it becomes essential to determine these operating limits, in particular with respect to voltage instability.

The objective of this paper is to give a description of the phenomena which contribute to voltage instability, to enumerate countermeasures, and to present in a (hopefully) unified and coherent way the computer analysis methods used or proposed.

The incidents experienced throughout the world and the threat of other blackouts have prompted significant research efforts among the power engineering community. The references given in this paper make up only a sample of the vast literature devoted to the subject. As “entry points” to this literature, let us point out:

1) early publications dealing with the subject [1]–[16];
2) a series of four seminars [17]–[20], which provided a forum for the presentation of research advances in the voltage stability area;
3) the reports of several CIGRE Task Forces [14], [21]–[24] and IEEE Working Groups [25]–[27] offering a compilation of techniques for analyzing and counteracting voltage instability;
II. VOLTAGE INSTABILITY PHENOMENA

A. Maximum Load Power

One of the primary causes of power system instability is the transmission of (large amounts of) power over long electrical distances. In voltage stability, attention is paid to power transfers between generation and load centers.

Let us first recall some fundamentals of the power transfer between a generator and a load. We use the simple model of Fig. 1, in which we consider for simplicity a purely reactive transmission impedance $jX$ and we assume that the synchronous generator behaves as a constant voltage source of magnitude $E$ (more realistic models will be discussed in the sequel).

Under balanced three-phase, steady-state sinusoidal conditions, system operation is described by the power flow or load flow equations [4], [28], [32]

$$P = -\frac{EV}{X} \sin \theta \quad (1)$$
$$Q = -\frac{V^2}{X} + \frac{EV}{X} \cos \theta \quad (2)$$

where $P$ (respectively $Q$) is the active (respectively reactive) power consumed by the load, $V$ the load bus voltage magnitude, and $\theta$ the phase angle difference between the load and the generator buses (see Fig. 1). Solving (1), (2) with respect to $V$ yields

$$V = \sqrt{\frac{E^2}{2} - QX \pm \sqrt{\frac{E^4}{4} - X^2P^2 - XE^2Q}}. \quad (3)$$

Fig. 2 shows how the terminal voltage $V$ changes with the load powers $P$, $Q$ (dimensionless variables are used in the figure). In “normal” conditions, the operating point lies on the upper part of the surface (corresponding to the solution with the plus sign in (3)), with $V$ close to $E$. Permanent operation on the lower surface, characterized by a lower voltage and higher current, is unacceptable.

The figure also confirms the existence of a maximum load power, well-known from circuit theory [33]. More precisely, the figure shows a set of maximum load power points, located on the “equator” of the surface (where the two solutions in (3) coalesce, i.e., the inner square root vanishes). The projection of this limit curve onto the $(P, Q)$ plane is the parabola shown in Fig. 2. In the $(P, Q)$ load power space, this parabola bounds the region where operation is feasible.

B. Nose Curves

In many reasonings (and industry practice) it is common to consider the curves which relate voltage to (active or reactive) power. Such curves, referred to as $PV$ (or $QV$) curves or nose curves are shown in Fig. 3, for our simple system. The curves depend on how $Q$ varies with $P$; in Fig. 3, a constant power factor, i.e., $Q = \tan \phi P$, has been assumed for each curve. This also corresponds to the solid lines in Fig. 2. Similarly, one may consider $PV$ curves under constant $Q$, or $QV$ curves under either constant power factor or constant $P$.

Simply stated, voltage instability results from the attempt to operate beyond maximum load power. This may result from a severe load increase or, more realistically, from a large disturbance that increases $X$ and/or decreases $E$ to the extent that the predisturbance load demand can no longer be satisfied. The latter scenario is illustrated in the next section, which offers a deeper look into a typical voltage instability mechanism, relating the latter to maximum load power as well as system theoretic concepts.

C. Long-Term Voltage Instability Illustrated on a Simple System

The following example, taken from [30], uses the simple system shown in Fig. 4(a). Bus 3 represents a distribution feeder. The power consumed at this bus may correspond to a large number of individual loads fed through medium voltage (MV) distribution lines, shunt capacitors, etc. We represent this aggregate load by the exponential model (widely used in large-scale stability studies)

$$P = P_o \left(\frac{V_o}{V^a}\right)^a \quad Q = Q_o \left(\frac{V_o}{V^b}\right)^b \quad (4)$$

where $V_o$ is a reference voltage and $P_o$ (resp. $Q_o$) is the active (resp. reactive) power consumed under this voltage.
Bus 3 receives its power from the transmission system through a transformer equipped with an automatic load tap changer (LTC). The objective of this device is to adjust the turn ratio of the transformer (in discrete steps) so as to keep the distribution voltage within some deadband $|V_0 - e V_o + e|$, in spite of voltage fluctuations on the transmission system. As in [30], we use "short-term" to refer to the time period unambiguously. As in [30], we use "short-term" to refer to the time period unambiguously. As in [30], we use "short-term" to refer to the time period unambiguously. As in [30], we use "short-term" to refer to the time period unambiguously.

Most of the load power is provided by a remote system (bus 1) through a rather long transmission. The remaining is supplied locally by the generator at bus 2. This generator is equipped with an automatic voltage regulator (AVR), in order to keep the voltage at bus 2 (almost) constant, and an overexcitation limiter (OEL), whose role is to prevent the rotor (or field) current from exceeding a specified thermal limit, in case the AVR imposes a sustained overexcitation.

We show and discuss hereafter two unstable responses obtained by simulating the tripping (at $t = 1$ s) of one circuit between buses 1 and 4.

**Case 1:** The system evolution, shown in Fig. 4(b)-(d), starts with electromechanical oscillations corresponding to swings of the generator rotor. These transients die out soon, indicating that the short-term dynamics of the synchronous generator are stable. Thus a short-term equilibrium is reached after about 10 seconds, with $V_4$ settling down close to 0.96 per unit.

The system response over the next minutes is a typical example of long-term dynamics, driven by the LTC and OEL.

The operation of the LTC starts after an initial delay of 20 s and continues at a rate of one tap change each 10 s. These changes further reduce the transmission-level voltage $V_4$.

The operation of the OEL can be seen from Fig. 4(c), showing the evolution of the generator field current. After the disturbance, this current jumps to 3 pu, which exceeds the limit, indicated by the dotted line. The OEL has an inverse-time characteristic, tolerating smaller overloads to last longer. Due to this delay, the OEL becomes active at $t = 70$ s. Before this time, the AVR controls the voltage at bus 2 and makes the field current further increase, in response to the first tap changes. This corresponds to a larger and larger active power drawn from the generator. After $t = 70$ s, each attempt to increase the field current is corrected by the OEL. With its field current kept (almost) constant, the machine behaves as a constant emf behind saturated synchronous reactance. The voltage at bus 2 is no longer controlled, and $V_4$ drops dramatically.

This drop goes on until $V_4$ reaches an unacceptably low value of 0.75 pu. At this point, the LTC has reached its limit and the voltage decline stops, since no other dynamic mechanism is involved.

The evolution of the distribution voltage $V_3$ and transformer ratio $r$ are shown in Fig. 4(d). Before the generator limitation, each tap movement produces the intended effect of rising $V_3$, bringing it back towards its deadband; on the other hand, after the limitation, the tap changes have negligible effect, then reverse effects.

A deeper look at the instability mechanism is provided by the $PV$ curves of Fig. 4(e), showing voltage $V_4$ as a function of the power $P$ transmitted to the load. All curves are drawn considering the short-term dynamics at equilibrium. The solid lines are the network $PV$ curves, similar to those shown in Fig. 3. On the other hand, through (4), the load power depends on $V_3$, which in turn depends on $V_4$ and $r$. The relationship between $P$ and $V_4$, for a given $r$, is the short-term load characteristics, shown with dotted lines in Fig. 4(e).

The system operates initially at point A. The disturbance causes the network characteristic to shrink from the rightmost to the middle curve. After the electromechanical oscillations, $r$ has not changed yet and the system operates at the short-term equilibrium point A'. Subsequent LTC operation brings the short-term equilibria to point B. At this point, the generator field current gets limited. Note the further decrease in maximum load power which results from the limitation of the generator reactive power. The system jumps to point B'. From there on, the LTC keeps on decreasing the tap until it finally hits its limit at point D.

The vertical dashed line in Fig. 4(e) is the long-term characteristic of the load. Due to the LTC action, a long-term equilibrium is such that $V_4 = V_o$, (ignoring the deadband $\pm e$), which means $P = P_o$. In other words, the LTC makes the load behave as constant power in the long term.

The nature of the instability is revealed by observing that this long-term load characteristic does not intersect the final network $PV$ curve. The maximum power that the final configuration can deliver is lower than what the LTC tends to restore. The system becomes unstable by loosing its long-term equilibrium.

**Case 2:** The system initial conditions are modified by increasing the local generator active production.

The response of the transmission voltage $V_4$ to the same disturbance is plotted in Fig. 5(a). In this case the generator field current gets limited at about $t = 100$ s. As the
circuit tripped at $t = 1\, \text{s}$

Fig. 4. Example system and simulation of Case 1 (all quantities in per unit) [30].

Fig. 5. Simulation of Case 2 (all quantities in per unit or radian) [50].

The $P \text{V}$ curves are shown in Fig. 5(c). As in the first case, the short-term equilibrium point follows the path $AA'$ (circuit tripping), $A'B$ (LTC operation), $BB'$ (generator limitation). Again, there is no long-term equilibrium with the generator limited. However, a major difference with respect to the previous case occurs when $r$ reaches the value 0.82: the short-term load characteristics does no longer intersect the network $PV$ curve, indicating that the system also looses short-term equilibrium, which corresponds to the above mentioned loss of synchronism.

D. Load Power Restoration

The previous example has shown a typical situation where the driving force of instability is the unsuccessful attempt of the LTC to restore the load voltage to its setpoint value and thereby the load power to its predisturbance level. The internal variable of this process if the transformer ratio $r$.

We mention hereafter two other well-identified load power restoration mechanisms.

1) Induction Motors [15], [28], [29], [41]–[43]: Induction motors are present in many industrial and commercial loads. When subject to a step drop in voltage, the motor active power $P$ first decreases as the square of the voltage $V$ (constant impedance behavior), then recovers close to its predisturbance value in the time frame of a second. The internal variable of this process is the rotor slip. In fact, a motor with constant mechanical torque and negligible stator losses restores to constant active power. Taking into account these losses and more realistic torque behaviors, there is a small steady-state dependence of $P$ with respect to $V$. The steady-state dependence of the reactive power $Q$ is a little more complex. $Q$ first decreases somewhat quadratically with $V$, reaches a minimum, and then increases up to the point where the motor stalls due to low voltage. In large three-phase industrial motors, the stalling voltage can be as low as 0.7 pu while in smaller appliances (or heavily loaded motors) it is higher.
Load restoration by induction motors may play a significant role in systems having a summer peak load, with a large amount of air conditioning [29, 44, 45].

2) Thermostatic Loads [13, 28, 29, 46, 47]: Another category of self-restoring load is the electrical heating controlled by thermostat.

A thermostat switches the heating resistor according to an on–off cycle, such that the energy produced over a cycle keeps the temperature within a deadband under the given weather condition. Following a decrease in voltage, the resistor power decreases as the square of the voltage. Over the next minutes, however, the on–off cycle changes progressively since the resistor has to stay on longer in order to produce the same energy. Considering the behavior of a large number of such devices over a small time interval, this increase in the on time appears as a recovery of the power to its predisturbance value. However, for a large enough voltage drop, the aggregate load power does not recover to its predisturbance value, owing to the fact that the heaters stay on permanently, thus giving a mere impedance load characteristic in the steady state. Also, it has been observed that the control cycle of the (older) bimetallic thermostat is itself influenced by the voltage, making load power restoration faster than what could be expected from thermal inertia [47].

Clearly, load self-restoration by thermostats is significant when analyzing the winter peak load of systems with a large amount of electrical heating. It is of concern mainly when the faster acting LTC's do not restore voltage, e.g., because they reach their limit.

E. Generic Models of Aggregate Loads

The load seen by a bulk power delivery transformer is an aggregate of many individual loads, fed through distribution lines and MV/LV transformers, compensated by switched capacitors, etc. The problem of modeling such an aggregate load is not easy to address. Indeed, while typical data can be obtained for every individual equipment [48], the real problem is to determine the composition of the load. The latter varies from one bus to another but also with the season, the time of the day, etc.

Although a detailed discussion of load modeling is out of scope of this paper, let us mention an approach proposed in the recent years [47, 49, 50, 52].

The response of aggregate loads to step decreases in voltage has been measured by several companies. The time evolution of power, over several minutes, recorded on the low voltage side of bulk power delivery transformers, is sketched in Fig. 6. The partial recovery originates from thermostatically controlled loads, downstream (nonmodeled) LTC's, voltage regulators, and voltage controlled capacitors, or even the consumers reaction to the reduced supplied power.

The exponential-type recovery shown in Fig. 6 may be captured by generic models, for instance

\[ P = z_P P_o \left( \frac{V}{V_o} \right)^{\alpha_t} \]

Table 1 Multitime-Scale Power System Model

<table>
<thead>
<tr>
<th>time scale</th>
<th>components</th>
</tr>
</thead>
<tbody>
<tr>
<td>long-term</td>
<td>LTCs, OELs, switched shunt compensation, aggregate load recovery, secondary</td>
</tr>
<tr>
<td></td>
<td>voltage and frequency control</td>
</tr>
<tr>
<td>short-term</td>
<td>generators, SVCs, AVR's and PSS's, turbines and governors, (equivalent)</td>
</tr>
<tr>
<td></td>
<td>induction motors, HVDC links</td>
</tr>
</tbody>
</table>

\[ \dot{z}_p = h_d(x, y, z_t, z_d) \]

\[ z_d(k + 1) = h_d(x, y, z_t, z_d) \]

where \( z_p \) is an internal state variable, obeying

\[ T_p \dot{z}_p = \left( \frac{V}{V_o} \right)^{\alpha_t} - z_P \left( \frac{V}{V_o} \right)^{\alpha_t} \]

with \( z_p^{\text{min}} < z_p < z_p^{\text{max}} \).

In some sense, the load obeys an exponential model which changes from the transient \( \alpha_t \) to the steady-state \( \alpha_s \) exponent. These exponents can be determined from the initial and final power drops (see Fig. 6) through

\[ \alpha_t \simeq \frac{\Delta P_o / P_o}{\Delta V / V_o} \quad \alpha_s \simeq \frac{\Delta P_o / P_o}{\Delta V / V_o} \]

while \( T_p \) can be obtained from a least-square fit of the time response (dotted line in Fig. 6). Similar relationships hold for the reactive power.

F. Time-Scale Decomposition Perspective

Table 1 enumerates the components, phenomena, controllers, and protecting devices which play a role in voltage stability, classified according to the time scale of the corresponding dynamics.

In stability studies, an instantaneous response is assumed for the network, according to the quasi-sinusoidal (or fundamental-frequency) assumption [32]. The network is thus described by the algebraic equations (7) derived from the Kirchhoff's current law at each bus, and involving the vector \( y \) of bus voltages magnitudes and phase angles.

*Obtained, for instance, by changing the transformer tap or by tripping one of two parallel transformers.
The short-term dynamics (8) last typically for several seconds following a disturbance.

The long-term components, acting typically over several minutes, may be represented either through discrete-time equations (9) (e.g., shunt compensation switching, LTC operation, the decision to switch the field current in an OEL, the changes in setpoints transmitted by the digital secondary voltage and frequency controllers) or through continuous-time equations (10) [e.g., the PID laws of these controllers, aggregate load models of the type (5), (6)].

There is a time decoupling between the short and long-term dynamics, which allows categorization of instabilities as explained hereafter [30], [53].

When a disturbance occurs in the system, the short-term dynamics are excited first. In this time frame the long-term variables \( z_c \) and \( z_d \) do not respond yet and can be considered as constant parameters in (8).

The short-term time scale is the time frame of angle instability, i.e., the loss of synchronism between generators. It is also the time frame of short-term voltage instability, which is linked to fast load recovery by induction motors and possibly HDVC systems.

In this time period, it may be difficult (and perhaps of academic interest only) to distinguish between angle and voltage instabilities. There are however some cases of "pure" voltage instability. Consider for instance the system of Fig. 1 where the load now consists of an induction motor.

1) Following a line outage, the maximum load power decreases as shown in Fig. 4(e). If it becomes smaller than the power the motor tends to restore, the latter stalls and the load voltage collapses. The system looses its short-term equilibrium;

2) A short-circuit near the motor causes the latter to decelerate. If the fault is not cleared fast enough, the motor is unable to reaccelerate and again, the load voltage collapses [44]. In this case, the long-lasting fault makes the system escape from the region of attraction of its postdisturbance equilibrium.

Let us assume that the system has survived the short-term period following the disturbance. From there on it is driven by the long-term dynamics (9), (10). Long-term voltage instability is associated with this slower dynamics. A typical example has been discussed in Section II-C, where the very cause of instability was the system loosing its long-term equilibrium.

The long-term evolution of \( z_c \) and \( z_d \) can be seen as changes of parameters imposed to the fast subsystem of (7) and (8). In many practical cases, the short-term dynamics respond in a stable way to these parameter changes. As long as this holds true, the analysis is considerably simplified by adopting the quasi-steady-state (QSS) approximation of long-term dynamics, which consists of replacing the short-term dynamic equation (8) by the corresponding equilibrium equation [30], [53]

\[
0 = f(x, y, z_c, z_d).
\] (11)

This approximation is valid along the whole system trajectory in Case 1 of Section II-C. However, it is possible that large changes in \( z_c \) and \( z_d \) eventually induce an instability of the short-term dynamics. In Case 2 of the same example, this takes on the form of the field current limited generator loosing synchronism. Motor stalling or oscillatory angle instability due to OEL's bypassing power system stabilizers (PSS) are other cases of induced short-term instability.

Obviously a real system evolution can be made more complex by, e.g., the action of protections that trip overloaded lines or limited generators operating at low voltages.

III. COUNTERMEASURES

Often prompted by emergency or blackout situations, a number of countermeasures have been adopted by power companies [23], [24]. Table 2 lists the most significant ones.

<table>
<thead>
<tr>
<th>Countermeasures Against Voltage Instability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System Reinforcement</strong></td>
</tr>
<tr>
<td>new transmission lines between generation and load centers</td>
</tr>
<tr>
<td>construction of generating stations: with low power factor, near load centers</td>
</tr>
<tr>
<td>series compensation shunt compensation, SVCs</td>
</tr>
<tr>
<td>devices and controllers line-drop compensation in AVRs</td>
</tr>
<tr>
<td>control of generator step-up transformers</td>
</tr>
<tr>
<td>automatic shunt compensation switching secondary voltage control</td>
</tr>
<tr>
<td>operational planning Voltage Security Assessment (VSA) commitment of out-of-merit units</td>
</tr>
<tr>
<td>real-time operation voltage profile and reactive reserve</td>
</tr>
<tr>
<td>monitoring and control</td>
</tr>
<tr>
<td>voltage security assessment</td>
</tr>
<tr>
<td>generation rescheduling</td>
</tr>
<tr>
<td>starting-up of gas turbines</td>
</tr>
<tr>
<td>preventive load shedding</td>
</tr>
<tr>
<td>system protection schemes</td>
</tr>
<tr>
<td>HVDC modulation</td>
</tr>
<tr>
<td>LTC emergency control</td>
</tr>
<tr>
<td>emergency load shedding</td>
</tr>
</tbody>
</table>

A. System Reinforcement

In the many cases where environmental, political, and other considerations leave little room for new transmission lines or power plants near populated load centers, other solutions must be sought.

The series compensation of transmission lines is a very effective way of decreasing transmission impedances and hence limiting the voltage drop over long distances [10]. However, these advantages must be balanced against considerations such as cost, more complex protections, and the possibility of a long-lasting bypass of the capacitors or subsynchronous resonance.

Shunt compensation has been the traditional way of providing the reactive power needed to maintain a good voltage profile. Capacitor banks are located near loads to improve the power factor and in the subtransmission systems to compensate for reactive losses. Excessive shunt compensation, however, has the drawback of bringing the maximum load power point closer to normal operating values, as can be seen from Fig. 3 for decreasing tan \( \phi \).
B. Devices and Controllers

Some devices and controllers contribute to maintaining transmission system voltages around their nominal values, after "normal" disturbances. Depending on their number, size, location, etc., they can also contribute to voltage stability.

Shunt compensation can be switched automatically, e.g., in response to low voltages. The speed of action is essential to counteract short-term voltage instability [27]; to this purpose, the more expensive but fast-responding SVC may be required [15]. For long-term instability, mechanically switched elements are often sufficient. In extra high voltage (EHV) systems, the shunt reactors used to prevent overvoltages may be tripped to counteract voltage instability [54].

The main control of transmission voltages is exerted by the AVR's installed on synchronous generators. A tighter control of the grid voltage near a power plant can be obtained by line-drop compensation, aimed at (partially) compensating the voltage drop in the step-up transformer impedance. Also, the on-load control of the step-up transformer ratio allows a wider range of network voltage variations. It can be easily seen from the simple example of Section II-A that a higher source voltage allows a higher power to be transmitted to the load.

On the other hand, the thermal constraints on field and armature currents impose hard limits on the voltage regulation capability of generators. These limits are usually presented to operators in the form of capability curves, which must be carefully checked against the real limiter settings. Although operating rules vary widely from one system to another, reactive reserves must be monitored and maintained (e.g., [55], [56]). Beside voltage profile aspects, the switching of shunt compensation by operators is also aimed at maintaining "dynamic" reactive power reserves, i.e., reserves on the fast-responding generators, synchronous condensers and SVC's.

Voltage control by AVR's is local by nature. As a consequence, after a disturbance, voltages at nongenerator buses may become unacceptable. Moreover, the required reactive power will be produced by the generators electrically closer to the disturbance and hence the reserves left may be unevenly distributed. The situation must be corrected by adjusting the AVR voltage setpoints. In many countries this is performed manually from a control center. A dedicated closed-loop control, referred to as secondary voltage control, has been implemented in France and Italy and planned in other European countries [57]–[60]. The objective is to control generators within zones, or regions, such that voltages at selected "pilot points" are kept close to setpoints, while making the reactive power production of each generator proportional to its reactive capability. The response time is in the order of 1 min, which is also the time frame of LTC operation.

This control provides some additional load power margin. In response to a load increase exceeding system capability, it tends to keep the voltage profile flat over a longer time interval, but it results in a sharper final decrease because all reactive reserves tend to be exhausted at the same time. Incidentally, it also makes network voltage a poorer indicator of an insecure situation.

To the author's knowledge, secondary voltage control has not been devised to face emergency situations, where a faster (and also coordinated) "boosting" of generator voltages is needed. The latter, if performed over a large enough area, could preserve generator reactive reserves by reducing network reactive losses and increasing the production of shunt compensation.

C. Operational Planning

Preventive evaluation of voltage security is mandatory in operational planning and is felt more and more as a necessity in the real-time environment of control centers. Computer methods used to this purpose are described in the next section.

In operational planning an important decision is the commitment of thermal units for security purposes. Deregulation is significantly changing the context in which such decisions are taken. For given transactions between the transmission system operator and the generation companies, security margins have to be checked, the congestions have to be identified, and the contracts adjusted accordingly. (For more details on market issues, please refer to the other papers in this issue.)

D. Real Time

Congestions due to voltage security may also appear in real-time operation. In this context, candidate preventive control may consist of rescheduling generation (shifting active power production, for instance, from cheap plants located far from a load center to closer but more expensive units) and/or starting up fast units (gas turbines, hydropower plant). In some cases, it may be decided to preventively shed load (in accordance with interruptible load contracts) if the system is still stable but security margins are deemed insufficient.

E. System Protection Schemes

The transmission open access and maximum profit environment will lead to operate power systems closer to their limits. In this context, the likely prevailing attitude will be to maintain the security margins needed to face the most credible contingencies and to rely on system protection schemes to contain more severe disturbances. By SPS we mean automatic, curative actions aimed at avoiding or containing instability. SPS's are cost effective solutions considering that preventive security has a price and severe disturbances have low probabilities of occurrence.

Although an SPS may integrate and coordinate several types of actions previously discussed, action on load is the ultimate countermeasure. This can be implemented indirectly through a modified control of LTC's or directly as load shedding.

The example of Section II-C has emphasized the role of LTC's in long-term voltage instability. Emergency control of LTC's can be achieved by LTC blocking, by bringing back...
the taps to predetermined positions, or by reducing LTC set-points (a 5% voltage drop is a typical figure). In the latter case, the sensitivity of load to voltage is exploited; from the viewpoint of customer voltage quality, although the distribution voltage remains low, it is on the average less sensitive to transmission system transients [27].

Many power systems have several LTC levels in cascade, for instance between successively EHV transmission, HV subtransmission, and MV distribution levels. The emergency actions have to be coordinated because the action on one LTC level affects the downstream voltages. One principle is to keep voltages low at the distribution level (to decrease the load power) and maintain normal (or even higher) voltages at subtransmission level (to keep low reactive power losses and get the most out of shunt compensation).

LTC emergency control slows down the system degradation but is counteracted by other load power restoration mechanisms. Also, it can be complex to implement due to the large number of distribution transformers to control.

Shedding a proper amount of load, at a proper place, within a proper time is the ultimate way of stopping voltage instability [45], [61]–[63].

Short-term voltage instability due to induction motors requires fast load shedding, i.e., response times as fast as 1.5 s. This is fast enough to prevent the stalling of large induction motors. Selective load shedding consists in tripping induction motor load, prone to short-term instability, in order to save the system. Thermal protection (set to tolerate the high starting current) cannot be relied upon to disconnect a motor; undervoltage protection is needed. Large industrial motors are equipped with such a protection but not the small motors used, e.g., in air conditioning, which may remain connected after stalling, drawing a large current and depressing voltages.

Let us quote a few aspects of load shedding against long-term instability, some of which being illustrated by the simple example of Section II-C [51], [64]–[69].

1) The primary objective is to restore a long-term equilibrium. With reference to Figs. 4(c) and 5(c), when some part of the load is shed, the long-term load characteristic (the vertical dashed line) is shifted to the left. For a large enough shift, the load and the postdisturbance network characteristics cross again, defining the new long-term operating point of the system.

2) As shown in Case 2, this action must take place before the system reaches a point where the short-term dynamics become unstable, i.e., where a true “collapse” or “system disruption” occurs.

3) Furthermore, load shedding must take place fast enough so that the restored long-term equilibrium is attracting (otherwise the system will not recover to this point but rather keep on plunging). It can be shown that, beyond some point in time, the longer one waits, the more one has to shed to save the system (and the higher the voltage jumps).

4) Another requirement may be to avoid a cascade of events such as overcurrent tripping of lines or undervoltage tripping of field-limited generators.

5) The location of the (active and reactive) load shedding matters when dealing with voltage instability. Some analytical methods can identify the most appropriate load buses, as outlined in Section IV-D. The farther the shedding from this location, the more one has to shed to save the system. This is significantly different from underfrequency load shedding, which only aims at maintaining a global active power balance.

All these aspects must be taken into account when devising load shedding SPS’s. The latter usually rely on low transmission voltages at major load buses, possibly combined with low reactive reserves on nearby generators. Several “blocks” of different magnitude can be shed, such as for underfrequency. Considering the grave consequences of load shedding, the smallest amount compatible with all the above requirements must be shed, which depends on the disturbance. The voltage thresholds must be set low enough to avoid undue action but high enough to meet the speed requirements 2) and 3) above. In voltage stability limited systems, severe disturbances may require to shed load a few seconds after the initiating disturbance, which invalidates a widespread belief that long-term voltage instability leaves more time to act. Even milder instability scenarios leave little time for an operator to react.

HVDC modulation can be also used to counteract short and long-term voltage instability, especially when the HVDC link exports active power from a weak ac area. The fast control of the dc power flow may help reducing both the active and reactive power required from the ac system [70].

IV. ANALYSIS METHODS

This section gives an overview of methods to analyze voltage instability scenarios and correspondingly assess system security. These methods are used at the various decision stages from planning to real-time. We have divided them into four categories, corresponding to the next four sections, respectively.

A. Contingency Analysis

Contingency analysis aims at analyzing the system response to large disturbances that may lead to instability and collapse. The system is considered secure if it can withstand each of a set of credible incidents, referred to as contingencies.

1) Criteria: Although criteria differ from one system to another, the following principles are widely accepted.

For long-term voltage stability analysis, the credible contingencies are outages of transmission and generation facilities; the sequence of events leading to such outages does not really matter. A well-known criterion is the $N - 1$ security according to which a system must be able to withstand any single transmission or generation outage. Multiple contingencies may also be considered; a typical example is the bus-bus fault cleared by tripping all equipments connected to the bar of concern. For
short-term voltage stability, the system response to short-circuits is investigated in addition to outages.

As already mentioned, postcontingency controls may contribute to stabilize the system. However, a common practice is to assess the system ability to survive the credible contingencies with the sole help of postcontingency controls that do not prevent load power restoration. For instance, shunt compensation switching or secondary voltage control will be considered, but not LTC blocking, LTC voltage reduction, nor load shedding, which impact on customers. Operator actions are not considered either, being deemed too slow.

Complementary to this, the adequacy of stronger controls is checked against more severe (i.e., less probable) disturbances, e.g., when designing SPS's.

2) Static Versus Time-Domain Methods: The example of Section II-C has emphasized that the main cause of long-term voltage instability is the loss of a long-term equilibrium. Basically, static methods focus on the existence of such an equilibrium; they rely on algebraic equations, derived from the equilibrium conditions of the dynamic model (7)–(10), which we will write in compact form as

\[ \varphi(u, p) = 0 \]  (12)

where \( u \) is a vector of state variables and \( p \) a vector of parameters. A detailed derivation and discussion is given in [30].

Static methods can be used in a rather wide range of systems and problems. Expectedly, they also have limitations. For instance, they cannot account for postcontingency controls that depend on the system time evolution. Generally speaking, time-domain methods are computationally more demanding but offer a higher modeling accuracy as well as the possibility to study more involved instability mechanisms, for instance when an equilibrium exists but is not attracting. Better interpretability of results (e.g., with respect to the sequence of events), possibility to obtain information on remedial actions, educational aspects, etc., are among the other advantages.

Voltage security is one aspect of dynamic security. It is often referred to as a separate class: of problems because, for a long time, dynamic security has been assimilated to transient (angle) stability only. Furthermore, the fact that static tools are often used to speed up computations should not lead to assimilating voltage security to static security in a confusion of means and ends.

3) On the Use of Load Flow Equations: Most often, the power system model at equilibrium (12) is approximated by the standard load flow equations. This approach is very popular because load flow programs are widely available and also used in static security analysis. Some limitations of this model are noteworthy.

1) Loads are represented as constant power. This is justified for a load controlled by an LTC, ignoring the dead-band of the latter. However, if the LTC hits a limit or in the absence of an LTC device, the long-term voltage dependence of the load should be taken into account. For instance, with the generic model (5), (6), an exponential model with exponent \( \alpha_d \) should be considered.

2) Generators are represented through either constant voltage or constant reactive power. Fig. 7 shows the voltage-reactive power characteristics of a generator at three levels of active power generation. Under AVR control, there is a voltage droop due to the use of a proportional controller (especially when the gain is low) [71]. Under field—and even more under armature—current limit, the reactive power varies with voltage. More importantly, the reactive power limit must be updated with the active power.

3) Instead of being left to the slack bus, any active power imbalance must be shared by generators according to governor, load frequency control, etc., effects.

Voltage-dependent loads and slightly more elaborate generator models can be incorporated to load flow calculations [9], [72]; an alternative is to consider QSS simulation is described later on in this section.

We discuss hereafter two static and two time-domain methods which can be used for contingency analysis.

4) Postcontingency Load Flow: For long-term voltage stability analysis, the simplest approach consists in computing the postcontingency long-term equilibrium. In unstable cases with no such equilibrium, any numerical method (usually the Newton algorithm) trying to solve these equations will diverge; this divergence can be used as an indication of instability.

Beside the already quoted limitations of static methods, this approach suffers from two drawbacks: 1) the divergence may result from purely numerical problems that do not (directly) relate to a physical instability; this is particularly true in load flow calculations where controls have to be adjusted and/or many generators switch under reactive power limit; 2) in truly unstable cases, we are left without information regarding the nature and location of the problem.

5) Modified Load Flows: Several modifications to the standard Newton method have been proposed to deal with cases of difficult convergence and/or absence of solution [65], [66], [73], [74]. In one of them, the sum \( E \) of squared mismatches is checked over successive iterations; when \( E \) is found to increase, a scaling factor \( \alpha < 1 \) is applied to the Newton correction. The value of \( \alpha \) is chosen to minimize \( E \) along the direction of the correction vector. The iterations are stopped when either: 1) all mismatches are negligible,
6) Multitime-Scale Simulation: Short-term voltage stability simulation requires the numerical integration of the differential-algebraic model (7), (8). Basically, this is the model used in transient (angle) stability studies, with proper account for load behavior. The numerical integration methods are the same [75].

Long-term voltage stability simulation requires the simulation of a whole model (7)–(10). This model is stiff, i.e., the dynamics of some part of the process are very fast compared to the interval over which the whole process is studied (compare for instance a damper time constant of a few hundredths of a second with the tap changing taking place over a few minutes). A number of approaches are in use for simulation of this stiff model [28], [76]–[80].

1) Compute the whole system response with the small time step size needed for accurate simulation of the short-term dynamics. This approach is simple since it merely requires to add the long-term models to a standard transient stability program. It is, however, a brute force approach resulting in large computing times and huge amounts of output points;

2) Increase the step size after the short-term dynamics have died out in order to subsequently filter out the fast transients that are not significant. A provision may be made to switch back to smaller step size upon detection of a fast instability. A criterion is to observe the rate of change of the fast variables.

3) Have the step size automatically adjusted to the system behavior, i.e., shorter when the fast dynamics are excited, longer when only slow transients are present.

In the last two approaches, the increase in step size requires a numerically stable integration method, otherwise numerical noise will grow up (even if, in the exact response, the fast transients have completely vanished). Implicit methods have much better numerical stability. A popular implicit integration formula is the trapezoidal rule; the latter, applied to (7), (8), takes on the form

\[ x_{n+1} = x_n + \frac{h}{2}(x_n + x_{n+1}) = c_n + \frac{h}{2}f(x_{n+1}, y_{n+1}) \]

0 = g(x_{n+1}, y_{n+1}).

An iterative predictor-corrector scheme is needed to solve these nonlinear equations. This scheme must also accommodate the increase in step size, without degradation of the convergence. Simultaneous integration using the Newton method appears as the most appropriate.

In the third approach, the step size can be controlled so as to:

1) keep the number of Newton iterations in between bounds [75];
2) maintain a constant integration accuracy; a well-known technique consists in estimating the local truncation error and adjusting the step size in order to keep this error estimate close to some tolerance [81].

In long-term voltage stability simulation of large systems, the numerous discrete-type devices (see Table 1) give rise to frequent discrete transitions. The latter are an obstacle to large increases in step size, because the transition times have to be identified with some reasonable accuracy, and too large time steps would eventually require some time consuming interpolation. Also, with a constant-accuracy control of integration, the fast dynamics induced by the above transitions may cause the step size to decrease very often. In practice, it can hardly be increased beyond several tenths of a second (which, however, yields a substantial gain). The above discontinuities can be avoided by using a continuous-time approximation of the LTC model or a generic model of the type (5), (6) to represent LTC-controlled loads.

In spite of computer power increase and algorithmic developments, multitime-scale simulation remains heavy in terms of computing times, data maintenance, and output processing.

7) QSS Long-Term Simulation: To speed up long-term voltage stability calculations, we have mentioned the use of larger time step sizes, which filter out the short-term transients. The next degree of approximation consists in simulating the long-term dynamics with the short-term dynamics replaced by their equilibrium equations. This yields the QSS approximation of the long-term dynamics, already mentioned in Section II-F. The idea is rather old [6], but in the recent years it has been developed either as one mode of operation of a dynamic simulation package [64], [76] or as a separate time-domain simulation program [82]–[84]. It also generalizes the empirical “step-by-step load flow” technique described in [26].

Formally, in the QSS model, (11) stems from the detailed short-term model (8) set to equilibrium; in practice, however, a reduced set of the above equations, involving a reduced state vector \( x \), is sufficient. For instance, three equations are sufficient to represent a synchronous generator, taking into account saturation, AVR, and speed droop effects [30].

Fig. 8 shows the output of a QSS time simulation8 and gives a detailed view of the sequence of computed points. Each dot represents a short-term equilibrium point, i.e., a solution of (7) and (11) with \( x_c \) and \( x_d \) fixed at their current values. The time step size \( h \) is in the order of 1–10 s in practice. The transitions from A to A′, B to B′, etc. come from the discrete dynamics (9) of LTC’s, OEL’s, etc. The short-term dynamics being neglected, there is no point in identifying very accurately the time of each transition; rather, the various discrete devices are checked “synchronously” at multiples of the time step \( h \) and switched once their internal delays are overstepped. The transitions from A′ to B, B′ to C, etc., correspond to the differential equations (10) and/or smooth variations of parameters with time (e.g., during load increase).

8The curve shows the unstable evolution of a 550-bus system [84], caused by a line outage at \( t = 2 \) s, followed by several shunt reactor tripings.
Points $A', B', \text{etc.}$, are obtained by solving (7) and (11) with respect to $x$ and $y$ using the Newton method, with a Jacobian updated and factorized as rarely as possible, e.g., after equipment trippings or OEL activation but not after tap changing or shunt compensation switching (unless otherwise required).

If there is no differential equation (10), the system evolution is a mere succession of short-term equilibria, and hence all points $A, A', B, B', \text{etc.}$ are computed as indicated above. To deal with differential equations, an explicit integration scheme is sufficient since the time step size $h$ is small compared to the time constants of (10). Moreover, a single-step integration method, with a partitioned solution scheme is best suited to the numerous discrete transitions.

QSS long-term simulation offers an interesting compromise between the efficiency of static methods and the advantages of time-domain methods. It is accurate enough for security analysis and adequate for real-time applications. On the other hand, as it focuses on long-term dynamics.

1) For severe disturbances, there is a risk of overlooking short-term instability. To handle these situations, an interesting solution is to couple the detailed short-term and the QSS long-term simulation, the former to check the system ability to survive the short-term period, the latter to check the long-term evolution.

2) QSS simulation cannot reproduce the final system collapse in cases where the short-term dynamics become unstable due to long-term instability. For instance, in the case of Fig. 5, QSS simulation stops when the $V_4$ voltage reaches 0.77 pu upon detection of a loss of short-term equilibrium. For voltage security assessment purposes, what happens to the system beyond this point is of limited interest, the scenario being clearly identified has long-term unstable.

**B. Loadability Limit Determination**

While contingency analysis focuses on a particular operating point, it may be also desirable to determine how far a system can move away from this operating point and still remain in a stable state. This type of analysis involves large but smooth deviations of parameters $p$, which we refer to as the system stress. The latter corresponds typically to load increases and/or generation reschedulings (e.g. within the context of a transaction) which stress the system by increasing power transfer over (relatively) long distances and/or by drawing on reactive power reserves.

**Table 3** Various Models Considered in Bifurcation Analysis

<table>
<thead>
<tr>
<th>type of analysis</th>
<th>Eq. (14)</th>
<th>Eq. (15)</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>short-term</td>
<td>(8)</td>
<td>(7)</td>
<td>$x_1$</td>
<td>$y_1$</td>
</tr>
<tr>
<td>voltage stability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>long-term</td>
<td>(8), (10) and</td>
<td>(7)</td>
<td>$x_2$</td>
<td>$y_2$</td>
</tr>
<tr>
<td>voltage stability</td>
<td>continuous-time</td>
<td>approxim. of (9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stability with</td>
<td>(10) and</td>
<td>(7,11)</td>
<td>$x_3$</td>
<td>$y_3$</td>
</tr>
<tr>
<td>QSS approxim.</td>
<td>continuous-time</td>
<td>approxim. of (9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A case of practical interest is when the stress is distributed over a set of buses, according to participation factors. This leads to parameter changes of the type

$$p = p^o + \mu d$$

where $\mu$ is a scaling factor and $d$ a "direction" of system stress. A loadability limit corresponds to the maximum value of $\mu$ such that the system remains stable.

The determination of loadability limits has prompted much interest for bifurcation analysis of power systems (let us quote nonexhaustively [85]–[91]). We briefly recall below some fundamentals of this analysis.

1) Voltage Instability and Bifurcations: The small-disturbance analysis of a power system relies on the differential-algebraic (DA) model

$$\dot{z} = h(z, v, p, \mu)$$

$$0 = g(z, v, p, \mu).$$

The correspondence between this and the previously discussed models is summarized in Table 3 [53].

Equations (14) and (15) are linearized into

$$\begin{bmatrix}
\Delta z \\
0
\end{bmatrix} =
\begin{bmatrix}
h_x & h_v \\
g_z & g_v
\end{bmatrix}
\begin{bmatrix}
\Delta z \\
\Delta v
\end{bmatrix}$$

where $h_z$ denotes the Jacobian matrix of $h$ with respect to $z$, etc. Assuming that $g_v$ is nonsingular and eliminating $\Delta v$ yields

$$\Delta z = H_z \Delta z$$

with

$$H_z = h_z - h_v g_v^{-1} g_z.$$  

Generally speaking, a bifurcation occurs when the qualitative structure of the system (i.e., the number of equilibria, their stability, etc.) changes for a small variation of the parameters. In the single-parameter family of DA problems, there are three generic bifurcations:

1) the saddle-node bifurcation (SNB), where two equilibrium coalesce and then disappear; at this point the Jacobian $H_z$ has a zero eigenvalue, i.e., is singular;

2) the (Poincaré-Andronov-)Hopf bifurcation, where there is emergence of oscillatory instability; at this
point, two complex conjugate eigenvalues of $H_x$ cross the imaginary axis;

3) the singularity-induced bifurcation, typical of DA systems, where $g_r$ is singular.

Since voltage stability relates to loss of equilibrium, the SNB points are primarily of interest. At such points, $H_x$ is singular; using Schur’s formula, it is easily shown that the “unreduced” Jacobian $J$ (see (16)) is also singular. Note that in $J$, both $h$ and $g$ (resp. $x$ and $v$) are handled in the same way. Thus, in general, if a system is described at equilibrium by equations of the type (12), the Jacobian $g'$ is singular at an SNB point.

This singularity condition, monitored through the determinant, the smallest singular value, or the eigenvalue closest to the origin has been extensively used in the literature [8], [26], [85], [92], [93]–[97].

When (12) is approximated by a load flow model, it can be easily shown (using again Schur’s formula) that the reduced Jacobian $J_{QV}$ of reactive power injections with respect to bus voltage magnitudes (under constant active power) is singular together with $\varphi$. The $J_{QV}$ matrix is at the heart of the approach proposed in [79] and [98], in which modal decomposition provides an algebraic interpretation of the eigenvectors relative to the (near) zero as well as other dominant eigenvalues.

At equilibrium, the sensitivities of a quantity $\eta(u, p)$ with respect to parameters $p$ are given by

$$S_{np} = \nabla_p \eta - \varphi_p^T (\varphi_u^T)^{-1} \nabla_u \eta$$

(19)

where $\nabla_p \eta$ (resp. $\nabla_u \eta$) is the gradient of $\eta$ with respect to $p$ (resp. $u$) and $\varphi_p$ the Jacobian of $\varphi$ with respect to $p$. Hence, the sensitivities tend to infinity as an SNB point is approached. Applications of this property have been considered in many papers, e.g., [8], [9], [97], [99]. Taking for $\eta$ the total reactive power generation and for $p$ the (vector of) bus reactive loads [83], [100]–[102] yields sensitivities which cover the whole system and merely require a factorization of the sparse $\varphi_u$ matrix and a single substitution.

2) Loadability Limits: Loadability limits can be pictured in the space of parameters $p$, as shown in Fig. 9(a). $\Sigma$ is the bifurcation surface, characterized by det $\varphi = 0$. When loads behave as constant power in the steady state, a convenient parameter space is the bus power space. For the simple system of Fig. 1, the $\Sigma$ surface in the $(P, Q)$ power space is the parabola shown in Fig. 2.

Starting from an initial operating point $O$ and stressing the system along direction $d$, the loadability limit is reached at point $L$. Tracing a component of $u$ as a function of $p$ yields Fig. 9(b), of which the traditional PV or QV curves are particular cases. This figure illustrates that the number of solutions $u$ of (12) may change with $p$. Only the upper branch of operating points (from $O$ to $L$) is generally of interest.

3) Effect of Generator Limits: As already shown in Section II-C, the generator reactive power limits introduce non-linearities which significantly reduce the loadability limit of a system.

This is illustrated with PV curves in Fig. 10(a). As the system is stressed with the generator under AVR control, the voltage decreases along the OA path. At the breaking point $A$, the generator reaches its limit. With the generator under OEL control, the voltage evolves along the AC path. Assuming a constant power load, the loadability limit is at point $C$, where the network and load characteristics are tangent and $\varphi_u$ is singular.

The switching of a generator from AVR to OEL control can be formulated through inequality constraints. Now, in the presence of such inequalities, a loadability limit may be reached at a point where the Jacobian $\varphi_u$ is nonsingular [30], [103]. For instance, in Fig. 10(b), no further stress is possible beyond point $A$; indeed, along path AB (assuming AVR control) the generator would be overloaded, while along path AC (assuming OEL control) the AVR would regain control.
Hence, the loadability limit is at point A, where the network and load characteristics are not tangent and $\varphi_u$ is nonsingular.

Linear-type indexes such as eigenvalues, singular values, and sensitivities undergo discontinuities when generators switch under limit. In the many cases where instability is precipitated by these switchings, linear indexes have rather poor anticipation capabilities. Linear analysis is more useful when carried out at the loadability limit provided by a nonlinear method, where it can provide an instability diagnosis (e.g., through eigenvector analysis, as explained in Section IV-D).

We now concentrate on four methods to obtain loadability limits.

4) Continuation methods [26], [99], [104]–[108]: The most natural way to determine a loadability limit is by computing the “solution path” of (12), (13) for varying $\mu$ (see Fig. 9(b)). In practice, only a small part of the lower solution branch is sought, the objective being to ascertain (e.g., visually) that the SNB point has been crossed.

Denoting by $n$ the dimension of both $\varphi$ and $u$, (12), (13) is seen as a set of $n$ equations in the $n + 1$ variables $u$ and $\mu$. A solution point is thus obtained by fixing one variable, called the “continuation parameter.” We restrict our description to the simple “local parametrization technique,” which works satisfactorily in most cases.

1) Far enough from the SNB point, $\mu$ is used as the “continuation parameter,” solving (12) and (13) repeatedly for increasing values of $\mu$. As the SNB point is approached, $\varphi_u$ may become ill-conditioned and beyond this point, (12) and (13) have no solution.

2) Therefore, once the Newton iterations diverge, the continuation parameter is changed from $\mu$ to one voltage magnitude (e.g., the lowest or the fastest decreasing) and (12) and (13) are solved for $\mu$ and the $n - 1$ other components of $u$ repeatedly for decreasing values of the fixed voltage. The crossing of the loadability limit is indicated by a decrease in $\mu$.

Generator limits are enforced as one progresses along the path. A predictor-corrector scheme can be used to speed up calculations. The secant predictor (or first-order extrapolation) does not require any extra computation. Step-length control has been also proposed to decrease the number of points on the low-curvature parts of the solution path.

5) Time Simulation Coupled with Sensitivity Analysis: Time simulation coupled with sensitivity analysis can also be used to determine loadability limits [63], [83].

The time response of the system to a ramp increase in $\mu$ is evaluated. The rate of increase of the stress must be as high as possible to save computing time but low enough so that the system can be considered to pass from one equilibrium to another.

Using a model which includes load power restoration, the SNB point is crossed without numerical difficulty. This crossing is characterized by one real eigenvalue of $\varphi_u$ going from negative to positive and sensitivities of the type (19) change sign through infinity. The latter can be computed at regular time intervals. QSS time simulation, stopped as soon as the sensitivities change sign, can be used to speed up calculations.

6) VQ Curves [14], [109]: A VQ curve expresses the relationship between the reactive power at a given bus and the voltage at that bus. It can be produced with a standard load flow program by adding at the bus of concern a fictitious generator with zero active power and recording its reactive production $Q$ as its voltage $V$ is varied. This technique is a particular continuation method, providing the loadability limit with respect to a single bus reactive load increase.

Fig. 11 shows VQ curves corresponding to a stable and an unstable situation, respectively.

In curve 1, point O is the system operating point, solution of (12), with the fictitious generator producing nothing. As its voltage is decreased, the generator consumes more and more reactive power, while after crossing the loadability limit $L$, the opposite happens. $Q_1$ is the reactive margin with respect to instability (by loss of equilibrium). Some planning criteria require the individual margins $Q_1$ relative to various buses to be above some threshold [56].

The fictitious generator, if properly located, may also help convergence in difficult cases. Assume, for instance, that convergence problems are met when seeking the solution point O, while with the fictitious generator holding its voltage at—say—one per unit, the solution O' is easily found. Point O can be reached by progressively decreasing the voltage of the generator until the latter produces no power.

Finally, curve 2 in Fig. 11 illustrates an unstable situation, with no system equilibrium. $Q_2$ is the margin to operability. The VQ curve can be used to determine the minimum shunt compensation needed to restore an operating point.

It must be emphasized that by varying the reactive power at a single bus, the system is driven to instability in a somewhat artificial way, very different from that imposed by a real power transfer.

7) Optimization Methods [72], [110]–[113]: Loadability limits can be computed as corresponding to the maximum value of $\mu$ such that (12) and (13) have a solution. Hence, the following optimization problem has to be solved:

$$\max \mu$$
subject to \ $\varphi(u, p^o + \mu d) = 0$.  \ (21)

The limit is computed directly, without determining the solution path between the base case and limit points. The optimization method is thus expected to be computationally more efficient than the continuation method. On the other hand, the latter is more attractive when the effect of controls acting along the solution path must be incorporated, or when it is of interest to obtain the solution path explicitly.

Defining the Lagrangian $L$ relative to (20) and (21), and setting to zero its derivatives with respect to $u$, $\mu$ and Lagrange multipliers $w$, we obtain the first-order optimality conditions:

$$\varphi_u^T w = 0$$
$$\varphi(u, p^o + \mu d) = 0$$
$$1 + w^T \varphi_{p^o} d = 0$$
Note that (22) expresses that $\varphi_u$ is singular, the vector of Lagrange multipliers being the left eigenvector of the zero eigenvalue. [94] used a formulation quite close to the above equations.

The nonlinear equations (22)–(24) can be efficiently solved by the Newton method. The corresponding Jacobian matrix is about twice as large as $\varphi_u$, but (being the Hessian of $L$) it is symmetric. It is also very sparse and can be stored and manipulated in blocked form. With proper initialization of $w$ the method converges quickly and reliably to the optimum.

The key point, however, is the handling of inequality constraints, mainly those relative to generator reactive power limits. As in any constrained optimization, the problem is to identify with reasonable effort which constraints are active at the solution. In the continuation method, generators are switched under limit along the solution path; in the optimization approach, this sequence of switchings is not known a priori. Reference [72] describes a procedure converging in a few steps to the set of limited generators, the corresponding equations (22)–(24) being solved at each step. As an alternative, the interior point method has been successfully applied to this problem. This approach, based on penalty "barrier" functions, can handle various types of inequalities. It requires to solve a sparse system of much greater size than (22)–(24); again, the use of efficient sparsity techniques is essential to preserve the computational advantages over the simpler continuation method.

In some applications the loadability limit can be obtained as the solution of an optimal power flow with a more "traditional" objective than (20). For instance the maximal power transfer from one area to another can be obtained as the solution of an economic dispatch problem with fictitious cheap (resp. expensive) production costs assigned to generators in the sending (resp. receiving) area. As another example, reactive power margins to stability can be computed, in unstable cases, by solving a sparse system of much greater size than (22)–(24).

There are basically two types of security limits. Given a direction of system stress and a list of contingencies:

1) postcontingency loadability limits (PCLL's) indicate how far the system can be stressed after the occurrence of each contingency;

2) secure operation limits (SOL's) indicate how far the system can be stressed prior to any contingency, such that it will remain stable after the contingencies [115], [116].

A PCLL provides a measure of the security margin left after a contingency. This notion is linked to the permanent character of contingencies (namely equipment outages). It can be computed by applying one of the methods described in Section IV-B to the postcontingency configuration of the system.

An SOL is easier to interpret insofar as it refers to precontingency parameters that operators can either observe (e.g., load increase) or control (e.g., generation rescheduling within the context of a transaction). Also, there is a clear separation between:

1) the precontingency configuration where operators and/or controllers react to the system stress, e.g., by switching shunt compensation, by adjusting generator voltages or transformer tap positions (on manual control), or through secondary voltage control; in practice, a load flow (or an optimal power flow) program is used to generate the precontingency operating points;

2) the postcontingency configuration, where only automatic controls are assumed to react; the system response to the contingency and these controls is analyzed by one of the methods described in Section IV-A.

For both types of limits, contingency filtering (or screening) is a key point for the success of real-time VSA. The objective is to quickly identify a group of contingencies containing the one(s) with the smallest security limit(s). Some methods have been proposed to avoid the huge computational effort of determining PCLL's for all contingencies (e.g., [55], [74], [117], [118]). The procedures for computing SOL's seem better suited to incorporate contingency filtering, as explained hereafter.

In the remainder of this section we describe one implementation of a combined SOL determination and contingency filtering procedure [116], [119].

Binary search is a simple and robust method to determine the SOL with respect to one contingency. It consists of building an interval $[S_l, S_u]$ of stress such that $S_l$ corresponds to a stable postcontingency evolution, $S_u$ to an unstable one, and $S_u - S_l$ is smaller than a specified tolerance $\Delta$. The search starts with $S_u$ set to $S_{\max}$, the maximum stress of interest and $S_l$ to a (possibly negative) lower bound of the sought limit. At each step, the interval is divided in two equal parts; if the midpoint is found stable (resp. unstable) it is taken as the new lower (resp. upper) bound. The procedure

\[^{10}\text{Also referred to as dichotomic search or bisection method.}\]
is illustrated in Fig. 12, with the arrows showing the sequence of calculations.

When the objective is to determine the SOL with respect to the most severe of a set of contingencies, it would be a waste of time to repeat the procedure of Fig. 12 for each contingency. It is more efficient to perform a “simultaneous” binary search (SBS) as outlined in Fig. 13. At a given step of the binary search, the various contingencies stemming from the previous step are simulated. If at least one of them is unstable, the stable ones can be discarded since their limits are higher than the current stress level; the search proceeds with the unstable ones only. As an interesting by-product, this procedure provides a lower bound of the security limit relative to each contingency; the smaller the limit, the more accurate this bound. It can be parallelized by distributing the contingency analyses over several computers.

One can also specify a level of stress below which all limits are calculated with the Δ accuracy and above which only lower bounds are provided. Indeed, it may be of interest to know others than the smallest limit, especially when the latter corresponds to a severe contingency with a low probability of occurrence, or a contingency for which instability does not affect a large part of the system.

For real-time and even operational planning applications, the contingency analysis methods compatible with the speed requirement appear to be the post-contingency load flow (possibly modified) [20] and the QSS time simulation [119].

We have already quoted the advantages of the latter in terms of accuracy, speed, and possibility of obtaining information in unstable cases. QSS time simulations themselves can be shortened by monitoring the total unrestored load power, i.e., the difference $C(t)$ between the total load power at time $t$ and its value at long-term equilibrium. Simply stated, in a stable case, $C(t)$ goes back to zero in the postcontingency period, while in an unstable case, it goes through a maximum and then decreases. The simulation can be stopped once it is clear that this maximum has been crossed.

When a large set of contingencies has to be processed by an SBS, a prefiltering (or screening) is essential. To this purpose, one may resort to a simplified computation. However, those controls with a negative impact on voltage stability (such as the switching of reactive power limited generators) must be taken into account. As for any contingency screening, a compromise must be found between simplicity, the overhead of too many false alarms and the risk of missing a harmful contingency.

Attention has been also paid to decreasing the number of simulations required to determine the SOL relative to a contingency, by extrapolating stability or instability indices obtained from time-domain simulations [119], [121]. This speed-up is more critical when multitime-scale simulation is used for contingency analysis. When dealing with several contingencies, this technique should be combined with SBS.

D. Preventive and Corrective Control

It is desirable to determine the best control actions to correct a weak situation. Preventive controls deal with actions to be taken in a precontingency situation in order to increase the security margin with respect to one (or several) “limiting” contingency( or contingencies). Corrective controls, on the other hand, deal with actions taken in a given postdisturbance configuration in order to restore system stability.

1) Preventive Control Identification Based on Eigenvector: As long as instability is linked to SNB, the identification of the most effective controls to correct a weak situation may rely on eigenvector information. The theoretical basis for using the (left) eigenvector is summarized hereafter [122].

Assume that the system described at equilibrium by (12) is stressed according to (13) up to reaching a loadability limit corresponding to the maximum value $\mu^*$ of the parameter $\mu$. It can be shown that the sensitivity of the margin $\mu^*$ with respect to a vector of parameters $q$ is given by [123]

$$S_{\mu^*q} = -\frac{\varphi_T^T w}{n^2 d} \quad \text{with} \quad n = \varphi_T^T w$$

(25)

where $\varphi_q$ (resp. $\varphi_p$) is the Jacobian of $\varphi$ with respect to $q$ (resp. $p$) and $w$ is the left eigenvector corresponding to the zero eigenvalue of $\varphi_q$. An interesting particular case is when the parameter space is the bus power space and $q$ is the vector of initial bus injections $p$. The above formula becomes

$$S_{\mu^*p} = -\frac{n}{n^2 d}$$

(26)

With reference to Fig. 9(a), it can be easily shown that $n$ is the normal vector to bifurcation surface $\Sigma$ at the SNB point

---

11Assuming that LTC's do not reach their limits.
L. Once point L and vector n are known, it is possible to build
the tangent hyperplane $\mathcal{H}$, which approximates linearly the
$\Sigma$ surface\(^\text{12}\) in the neighborhood of L.

In preventive mode, the sensitivities (25) allow one to de-
termine the change in margin induced by a given change in
parameters or conversely, to find the amount of parameter
variation required to increase the margin by a given amount.
In particular, (26) can be used to determine suitable locations
where generation should be increased or load decreased.

Incidentally, the above properties have inspired research
work on the determination of the closest bifurcation point in
the parameter space [in Fig. 9(a), the point of $\Sigma$ closest to the
operating point $O$] [124] or the optimization of the margin
with respect to parameters [125]. Note that the pure criterion
of minimum Euclidian distance in the bus power space may
lead to unrealistic load increase patterns.

Among the loadability limit determination methods de-
scribed in Section IV-B, the optimization approach automatically
provides $w$ in the form of Lagrange multipliers, as can be
seen from (22). In other approaches $w$ has to be computed
separately. This computation is not demanding as far as
one focuses on the zero eigenvalue only, in practice on the
dominant real eigenvalue. Iterative methods based on the Si-
multaneous Iterations algorithm [28], [98] can converge to
the eigenvalue closest to an estimate $\lambda$ (zero or a small pos-
itive number in our case) while avoiding convergence prob-
dleas due to the presence of several other eigenvalues close
to $\lambda$. At each iteration, a linear system with the sparse matrix
$J - \lambda I$ is solved.

2) Corrective Control Identification from Eigen-
vector: Coming back again to Fig. 9(a), assume now that
the system operates initially at the long-term equilibrium $O'$,
which is inside the predisturbance bifurcation surface $\Sigma'$.
In an unstable case, a severe disturbance causes this surface
to shrink from $\Sigma'$ to $\Sigma$, so that point $O'$ is left outside the
new feasible region. As already mentioned, the objective
of corrective actions such as load shedding is to restore an
equilibrium, i.e., to move $O'$ inside the new feasible region.
If one point of $\Sigma$ can be obtained, the tangent hyperplane ap-
proximation allows to identify the most effective corrective
controls as corresponding to the shortest distance between $O'$ and $\mathcal{H}$.

In static approaches, one can obtain one point of $\Sigma$ using a
modified load flow [65], [66] or a continuation method [126].
When using (QSS) time simulation, this point can be identified
along the unstable system trajectory, through a change in
sign of sensitivities, as explained in Section IV-B [64], [69],
[83].

A minimum distance in the Euclidian sense may be used to
determine the optimal corrective control [65], [66]. In prac-
tice, one corrective control may involve several parameter
changes (e.g., both the active and reactive powers are de-
creased when shedding load); it is thus appropriate to con-
sider directions of parameter changes, such as $c_1$ or $c_2$ in
Fig. 9(a). In this example, $c_1$ is more efficient than $c_2$ since

\(^{12}\)Equation (26) can be obtained from simple geometrical considerations
using $\hat{\Sigma}$.

VAN CUTSEM: VOLTAGE INSTABILITY

3) Preventive Control Based on Optimal Power Flow: In
Section IV-B we presented optimization methods as one way of
computing loadability limits. The optimization problem,
defined by (20) and (21), had a single degree of freedom.
Optimization methods also allow to incorporate adjustable
controls and compute the larger loadability limit which can
be achieved with these additional degrees of freedom [128],
[129]. For instance, generator powers or voltages may be let
free to vary within specified ranges.

4) Corrective Control Based on Optimal Power Flow: As
recalled above, the primary objective of corrective control is
to bring back point $O'$ into the feasible region of operation,
in some optimal manner. This problem, often referred to as
solvability restoration, can be formulated as an optimization
problem: minimize the amount of corrective actions, subject
to the system equilibrium equations (12) and the inequalities
corresponding to—at least—generator reactive power limits.
For instance, in order to determine the minimal load shed-
ing, the objective function can be taken as $\sum_i f_i P_i$, where
$P_i$ is the load active power and $f_i$ the (unknown) fraction
of the load shed at the $i$-th bus (the same fraction applies to
the reactive load). Upper bounds can be specified on $f_i$, e.g.,
to obey load interruptibility contracts. The interior point
method has been successfully applied to this particular op-
timal power flow problem [128], [129].

This approach identifies the most appropriate controls and
determines their optimal variations in a single step. It also
offers the possibility of incorporating operating constraints
in addition to generator reactive limits. On the other hand,
rely on a load flow model and should be complemented
with a verification of the control timing aspect.

V. PERSPECTIVES

Over the past decade, significant progress has been made
in the understanding of voltage instability mechanisms, the
implementation of countermeasures, and the development of
computer methods. This paper has described (nonex-
ahustively) some of the current practices as well as some
promising approaches.
By way of conclusion, we would like to mention some topics which, in our opinion, deserve particular attention.

1) Load behavior is at the heart of voltage instability. As far as loads restore to constant power due to LTC action, their short-term behavior has a rather limited impact; however, this is no longer true when the problem is to design a controller or a system protection scheme aimed at acting quickly after a disturbance or during the load recovery phase. Similarly, security limits which involve a regional load increase are somewhat affected by the uncertainty on the participation of individual loads. This leads to operate the system with some security margin from the computed limit (typically a few percents of the latter). More work will certainly be done in the area of load model identification from field measurements. Nevertheless, the aggregate loads seen from bulk power delivery transformers are likely to remain the most uncertain power system components. Statistical evaluations of the impact of this uncertainty are needed, for both preventive security analysis and corrective control design.

2) The next challenge is the development and implementation of a true VSA in the real-time environment of control centers. To the author's knowledge there are few energy management systems including this function. Let us quote more specifically: the efficient determination of security regions (e.g., in the context of multiple transactions), the handling of multiple contingencies in preventive security enhancement, the real-time modeling of external and lower voltage-level systems, the development of indexes to measure the severity of instability (e.g., with respect to the extent of the affected area), a more refined evaluation of reactive power reserves taking contingencies into account, etc.

3) Complementary to preventive security analysis, there is a need for methods allowing a better and more automatic design of system protection schemes, in particular undervoltage load shedding. Closed-loop control is needed to face system modeling uncertainties. Methods are needed to optimize the protection parameters over a large set of scenarios. Besides the traditional protection scheme gathering local information, can we think of a protection based on real and reactive currents, Revue Générale de l'Electricité, vol. 8, pp. 1-17, 1986.

4) While proven methods are available to speed up the analysis of long-term voltage stability, the counterpart in short-term voltage stability is not so much developed. Attention should be also paid to situations where a higher stress or a higher disturbance severity makes the instability change from the long-term to the short-term type.

5) Very promising results have been obtained in the field of Automatic Learning methods [130]-[132], aimed at extracting from large statistical data bases of scenarios a higher-level information on the system behavior.

ACKNOWLEDGMENT

The author would like to thank C. Vournas of the National Technical University of Athens, Greece, whose valuable discussions benefited this paper.

REFERENCES


Thierry Van Cutsem (Member, IEEE) was born in Charleroi, Belgium, in 1956. He graduated in electrical and mechanical engineering from the University of Liège in 1979, and he obtained the Agrégé de l'Enseignement Supérieur (Ph.D.) degree in 1984 from the same institution.

Since 1980 he has been with the FNRS (Belgian National Fund for Scientific Research), where is now Research Director. He is also lecturer at the University of Liège, Belgium. His research interests are in power system dynamics, control, and stability, numerical simulation, and security analysis. He has developed a software for voltage stability and security analysis which is used by several power companies. He is co-author of the monograph Voltage Stability of Electric Power Systems (Kluwer, 1998).

Dr. Van Cutsem is a member of several IEEE and CIGRE working groups.