View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Liverpool Repository

FACTS Devices Allocation via Sparse Optimization

Chao Duan, *Student Member, IEEE*, Wanliang Fang, L. Jiang, *Member, IEEE*, Shuanbao Niu

***Abstract*—Although there are vast potential locations to install FACTS devices in a power system, the actual installation number is very limited due to economical consideration. Therefore the allocation strategy exhibits strong sparsity. This paper formulates FACTS device allocation problem as a general sparsity-constrained OPF problem and employs $L_q(0 < q \leq 1)$ norms to enforce sparsity on FACTS devices setting values to achieve solutions with desirable device numbers and sites. An algorithm based on alternating direction method of multipliers is proposed to solve the sparsity-constrained OPF problem. The algorithm exploits the separability structure and decomposes the original problem into an NLP subproblem, an L_q regularization subproblem, and a simple dual variable update step. The NLP subproblem is solved by the interior point method. The L_q regularization subproblem has a closed-form solution expressed by shrinkage-threholding operators. The convergence of the proposed method is theoretically analyzed and discussed. The proposed method is successfully tested on allocation of SVC, TCSC and TCPS on IEEE 30-, 118- and 300-bus systems. Case studies are presented and discussed for both single-type and multiple-type FACTS devices allocation problems, which demonstrates the effectiveness and efficiency of the proposed formulation and algorithm.**

***Index Terms*—Flexible AC transmission system, sparse optimization, optimal power flow, L_q norm, alternating direction method of multipliers.**

I. Introduction

FACTS devices can be utilized to increase transmission capability and improve stability in modern power systems. In order to maximize the benefits of installing FACTS devices, their types, location, capacity, and even initial settings, should be systematically determined. Usually this problem is called a FACTS devices allocation problem and has attracted much attention in the past two decades [1]–[5]. Various formulations and algorithms have been put forward to deal with this problem. Different objective functions are proposed from the perspectives of system economy and/or security, such as investment costs [1], [2], transmission losses [3], generator fuel costs [4], voltage stability index [5], voltage profile [2] and system loadability [6].

In fact, this problem is theoretically a mixed integer nonlinear programming (MINP) problem, and still there are no general and effective mathematical techniques to solve such problem, especially when the scale of the problem is large. To handle the difficulty of MINP, sensitivity analysis [7], intelligent optimization algorithms [8], [9] and mixed integer linear programming (MILP) have been extensively investigated in previous literatures [1], [10], [11].

The basic idea of sensitivity analysis methods is to find some indicators to determine the most critical transmission lines or buses for installing the FACTS devices [7]. Sensitivity analysis methods have their advantages over other optimization-based methods in computing efficiency. However, their computation accuracy is partly lost, because the nonlinearity of the power flow model is neglected. Moreover, they can not simultaneously optimize the device number, location as well as initial settings.

Intelligent optimization algorithms have their merits in dealing with discrete variables and finding the global optimal solution, so they have been widely applied to FACTS devices allocation problems [8], [9]. Unfortunately the demerits of this type of algorithms are their high computational burden.

MILP based algorithms either relax or approximate the original nonlinear formulation to a linear model and they can be classified into two groups: the relaxation group and the approximation group. In the relaxation group, decomposition techniques are employed successively to reformulate the original MINP into MILP [1], [10], [11]. However, recursively solving MILP is too time-consuming for large-scale systems. In the approximation group, the nonlinear power flow is approximated to simpler models such as the DC model [12] and the simplified LFB (line flowed based) model [2]. Though algorithms in the approximation group have relatively high computational efficiency, the approximation nature of the problem formulation makes those algorithms merely suitable for initial analysis in power system planning, and their results need to be further refined by a full AC power flow model.

Problems in various science and engineering fields motivate the need for a sparse solution [13]. In 2012, R. A. Jabr et al. applied sparse optimization for the first time in power system VAR planning [14] , where L_1 regularization was combined with successive conic programming to achieve a sparse solution vector, i.e. a VAR allocation strategy with minimum installation sites.

In fact, FACTS devices allocation problems possess following features: 1) large-scale mixed integer nonlinear programming problems; 2) a large number of transmission lines and buses which implies a large quantity of potential locations to install FACTS devices; 3) limited actual number of FACTS devices to be installed due to economical consideration. Considering those features, in this paper, FACTS devices allocation problems are considered as a sparse optimization problem, by introducing an extra constraint, i.e. the solution vectors must be sparse [15].

A new algorithm is proposed to solve these sparse optimization problems, which consists of the following four parts: 1)

This work was supported in part by the State Grid Corporation of China Research Project (No. YHGC-201206-XBW)

Chao Duan and Wanliang Fang are with the Department of Electrical Engineering, Xian Jiaotong University, Xian 710049, China. Chao Duan is also with Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, U.K. Email: duanchao@stu.xjtu.edu.cn; eewlfang@mail.xjtu. edu.cn.

L. Jiang is with Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, U.K. Email: ljiang@liv.ac.uk.

Shuanbao Niu is with the Northwest China Grid Company, Xian 710048, China. Email: 149991@qq.com.

sparsity-inducing norms [13], 2) alternating direction method of multipliers (ADMM) [16], 3) interior point method (IPM) [17], [18] and 4) shrinkage-thresholding operators (STO) [19]–[21], and named as ADMM-IPM-STO. The objective function (system loadability) is regularized by several sparsity-inducing norms. $L_{1/2}$ and $L_{2/3}$, recently developed in [20] and [21], are firstly introduced in power system optimization in this paper. Additionally, L_1 norm is also employed in the numerical experiments. The features of different sparsity-inducing norms is discussed. The sparse optimization problem is decomposed into two sub-problems by ADMM, namely a nonlinear programming (NLP) sub-problem and an L_q regularization sub-problem. Then the NLP sub-problem is solved by IPM, while the L_q regularization sub-problem has a closed-form analytic solution expressed by STO. The theoretical convergence of the proposed method is analyzed. A weak result that guarantees the optimality of the practical solutions of all $q \in (0, 1]$ and a strong result that substantiates the convergence for $q = 1$ are obtained.

Within above framework, the state-of-art method for sparse optimization is combined with one of the most successful methods for optimal power flow to form a novel FACTS device allocation algorithm. The numbers, locations, initial settings and even types of FACTS devices are simultaneously determined by this algorithm. Its computation precision is relatively higher than that of DC or LFB based algorithms, because the exact AC flow model is adopted. On the other hand, the proposed algorithm has relatively higher computational efficiency than intelligence optimization algorithms which also use AC flow model. Furthermore, this algorithm is adaptive to various types of FACTS allocation problems and possesses the flexibility to use different sparsity-inducing norms to achieve desirable sparse features.

Case studies are carried out on IEEE test systems from 9 buses to 300 buses, respectively, which demonstrates validity and above mentioned advantages of the proposed algorithm.

The rest of this paper is organized as follows. Notations used in this paper are defined in section II. The mathematical formulation of general sparse optimization problem and the ADMM-IPM-STO algorithm, along with the convergence analysis of ADMM-IPM-STO are introduced in section III. The sparse optimization model for general FACTS devices allocation problems is presented in Section IV. Case studies are reported in Section V. Finally, Section VI draws conclusions and gives suggestions on future research.

II. Notation

The imaginary unit is denoted by j. Boldface lower case letter $\boldsymbol{a}$ represents a real vector and its i^{th} element is denoted by a_i. Hatted boldface lower case letter $\hat{\boldsymbol{b}}$ represents a complex vector with its i^{th} element denoted by $\hat{b}_i$. The set of real and complex n-vectors are denoted by $\mathbf{R}^n$ and $\mathbf{C}^n$.The conjugate of a complex number $\hat{b}$ is denoted by $\hat{b}^c$. $\text{Re}(\hat{b})$ and $\text{Im}(\hat{b})$ denote the real and imaginary part of $\hat{b}$. We denote the gradient and hessian of the function $f(\boldsymbol{x})$ in the point $\boldsymbol{x}^*$ as $\nabla f(\boldsymbol{x}^*)$ and $\nabla^2 f(\boldsymbol{x}^*)$, respectively. $\nabla^2 f(\boldsymbol{x}^*) \succcurlyeq \mathbf{0}$ denotes the hessian matrix of f at $\boldsymbol{x}^*$ is semidefinite. The directional derivative of $f(\boldsymbol{x})$ in the point $\boldsymbol{x}^*$ toward direction $\boldsymbol{d}$ is denoted by $f'(\boldsymbol{x}^*; \boldsymbol{d})$. Vector and scalar sequences are denoted with superscript like $\boldsymbol{x}^k$ and α^k. The δ-neighborhood of vector $\boldsymbol{x} \in \mathbf{R}^n$ is denoted as $B(\boldsymbol{x}, \delta) = \{\boldsymbol{y} \in \mathbf{R}^n | \ \|\boldsymbol{y} - \boldsymbol{x}\|_2 < \delta\}$. We use parentheses to construct vectors from comma separated lists as $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_k^T]^T$.

In addition, the following special symbols are used in our problem formulation:

n_b	number of buses
n_g	number of generators
n_l	number of lines
$\hat{y}$	series admittance of a line
jb	shunt admittance of a line
$\hat{v}$	complex voltage at a bus
e	real part of the voltage at a bus
f	imaginary part of the voltage at a bus
$\hat{i}_g$	complex current injection of a generator
a_g	real part of the current injection of a generator
b_g	imaginary part of the current injection of a generator
b_{sh}	susceptance of a shunt compensator
κ	compensation rate of a series compensator
φ	angle of a phase shifter
τ	ratio of a controllable transformer

III. Formulation and Algorithm for General Sparse Optimization

A. Sparse Optimization Problems

A standard nonlinear optimization problem usually is expressed as

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in S} f(\boldsymbol{x}, \boldsymbol{u}) \tag{1}$$

where $f(\cdot)$ is a nonlinear scalar objective function, $x \in \mathbf{R}^m$ and $u \in \mathbf{R}^n$ are two sub-decision vectors. S is the feasible set of the decision variables, and it will be specified by equality and inequality constraints later on.

In some applications, e.g. FACTS devices allocation problems, decision vector $\boldsymbol{u}$ is expected to be very sparse after optimization. This expectation can be considered an extra constraint in above original optimization problem (1). This kind of problem is named the sparsity-constrained optimization problem which is intuitively transformed into the so-called L_0 regularization problem [22]:

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in S} f(\boldsymbol{x}, \boldsymbol{u}) + \lambda \|\boldsymbol{u}\|_0 \tag{2}$$

where $\|\cdot\|_0$, called L_0 norm, is the number of nonzero components of $\boldsymbol{u}$. The non-negative parameter λ, given by decision-makers, balances the two objective terms. Obviously, the larger λ is, the sparser induced $\boldsymbol{u}$ will be. However, L_0 regularization problem is NP hard [23] due to the discrete nature of L_0 norm. Actually, it can be seen from (4) that discrete L_0 norm is the limit of the q power of continuous L_q norm when q approaches 0. A natural way to overcome the difficulty in solving the L_0 regularization problem is to consider the so-called L_q regularization problem

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in S} f(\boldsymbol{x}, \boldsymbol{u}) + \lambda \|\boldsymbol{u}\|_q^q \tag{3}$$

where

$$\|\boldsymbol{u}\|_q = (\sum_{i=1}^{n} |u_i|^q)^{1/q} \quad (0 < q \leq 1) \tag{4}$$

According to the definition (4), L_q norms, to some extent, are approximations to L_0 norm. As q approaches 0, L_q norm mainly exerts penalty on the number of nonzero components of the solution vector. Comparatively, as q reaches 1, L_q norm penalizes the sum of absolute values of the solution vector components. When q lies somewhere between 0 and 1, L_q norm exerts penalty on both the number of nonzero components and the sum of absolute values to a certain degree. In other words, for a certain λ, the sparsity of the solution vector to the L_q regularization problem increases as q decreases. The special importance of $L_{1/2}$ regularization is highlighted in [24] by showing the representativeness of $L_{1/2}$ regularization among all L_q regularizations. This work basically reveals that the sparsity of the L_q solution significantly increases as q decreases when $1/2 < q \leq 1$ but is insignificantly affected by q when $0 < q \leq 1/2$. In addition, thresholding representation theories have been developed for L_q regularization problems when $q = 1/2$, $2/3$ and 1 [19]–[21], which leads to efficient algorithms for solving those L_q regularization problems. Therefore, $L_{1/2}$, $L_{2/3}$ and L_1 norms are so far the best sparsity-inducing norms to obtain desirable sparse solutions.

B. ADMM-IPM-STO Algorithm for Sparse Optimization

Consider the L_q regularization problem (3). The objective function consists of two terms and the second term is a continuous, non-smooth, non-lipschitz function of $\boldsymbol{u}$. Conventional joint minimization methods are incapable to tackle this problem.

ADMM is one of the state-of-art methods for sparse optimization. A comprehensive account of ADMM appears in [16] and its applications in sparse optimization are reported in [25]. ADMM is actually a version of the method of multipliers where Gauss-Seidel iterations are used to separately minimize two terms in the objective function instead of conventional joint minimization. ADMM utilizes the separability structure of the objective in (3) and decomposes problem (3) into two simpler sub-problems.

In order to make the objective separable, first an auxiliary vector $\boldsymbol{v}$ and an auxiliary equality constraint are introduced, and then problem (3) can be equivalently transformed into

$$\begin{aligned} &\min_{(\boldsymbol{x},\boldsymbol{u})\in S, \boldsymbol{v}\in\mathbf{R}^n} f(\boldsymbol{x},\boldsymbol{u}) + \lambda\|\boldsymbol{v}\|_q^q \\ &s.t. \quad \boldsymbol{u} - \boldsymbol{v} = \mathbf{0} \end{aligned} \tag{5}$$

The augmented Lagrangian function with respect to the auxiliary equality constraint is given by

$$L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y}) = f(\boldsymbol{x},\boldsymbol{u}) + \lambda\|\boldsymbol{v}\|_q^q + \boldsymbol{y}^T(\boldsymbol{u}-\boldsymbol{v}) + \rho/2\|\boldsymbol{u}-\boldsymbol{v}\|_2^2 \tag{6}$$

where $\boldsymbol{y}$ is an n dimensional Lagrangian multiplier vector related to the auxiliary equality constraint, ρ is a positive penalty parameter.

ADMM-IPM-STO algorithm consists of the following iterations:

$$(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}) := arg \min_{(\boldsymbol{x},\boldsymbol{u})\in S} L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}^k,\boldsymbol{y}^k) \tag{7}$$

$$\boldsymbol{v}^{k+1} := arg \min_{\boldsymbol{v}\in\mathbf{R}^n} L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v},\boldsymbol{y}^k) \tag{8}$$

$$\boldsymbol{y}^{k+1} := \boldsymbol{y}^k + \rho(\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1}) \tag{9}$$

Note that problem (7) is a conventional continuous nonlinear optimization problem whose decision vectors are only $\boldsymbol{x}$ and $\boldsymbol{u}$. Various methods have been proposed to solve this NLP problem, among which IPM has experienced great success. Specially, IPM has almost become a standard method to solve optimal power flow problems in recent years. Hence, IPM is chosen to solve the NLP sub-problem. When problem (7) solved, $\boldsymbol{x}^{k+1}$ and $\boldsymbol{u}^{k+1}$ are obtained. Problem (8) thus can be equivalently stated as

$$\boldsymbol{v}^{k+1} := arg \min_{\boldsymbol{v}\in\mathbf{R}^n} (\lambda/\rho)\|\boldsymbol{v}\|_q^q + (1/2)\|\boldsymbol{u}^{k+1} - \boldsymbol{v} + (1/\rho)\boldsymbol{y}^k\|_2^2 \tag{10}$$

So far, there is no general theoretical understanding and efficient algorithms to arbitrary $q \in (0, 1]$, because L_q norm in (10) is non-convex, nonsmooth and non-Lipschitz. Fortunately, at special points of $q = 1/2$, $2/3$, and 1, closed-form analytic solutions have been established in [19]–[21] which give global optimum for (10). They all can be expressed as shrinkage-thresholding operators. Though the exact forms vary with different sparsity-inducing norms, they are uniformly given by

$$\boldsymbol{v}^{k+1} = \text{Shrink}\left(\boldsymbol{u}^{k+1} + (1/\rho)\boldsymbol{y}^k, \lambda/\rho\right) \tag{11}$$

Specific results for $L_{1/2}$, $L_{2/3}$ and L_1 norms are presented in the Appendix A. The computation burden of obtaining $\boldsymbol{v}^{k+1}$ can almost be omitted since above closed-form solutions involve no iterations. Then according to the method of multipliers, dual variable $\boldsymbol{y}^k$ needs to be updated to $\boldsymbol{y}^{k+1}$ as in (9). The iteration terminates when the primal error (12) and dual error (13) are both small enough.

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1}\|_2 < \epsilon_{primal} \tag{12}$$

$$\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\|_2 < \epsilon_{dual} \tag{13}$$

C. Convergence Analysis of ADMM-IPM-STO

The convergence of ADMM is well-established for convex problems in [16]. So far, there is no global convergence result of ADMM for general non-convex optimization problems. Even so, ADMM has been extensively applied to non-convex problems [26], [27], including OPF problems [28], [29] and shown robust performance in practice. The convergence analysis of an optimization algorithm can be divided into two questions. First, whether does the algorithm generate a limit point? Second, whether is the limit point an optimum? Both questions are of crucial importance to substantiate the algorithm. In engineering application, the second question is to some extent more important than the first one because emergence of a limit point can be directly observed through numerical computation while optimality is not obviously available. Since the proposed problem formulation is non-convex, and even

non-lipschitz, the above two questions are in doubt. In this part, we establish a weak result for problem (5) with $0 < q \leq 1$ to answer the second question and a strong result with $q = 1$ to partly answer the first one. In the weak result, we basically adopt the same approach as in [26] and [27], whereas the strong result is essentially based on [16].

Because $|x_i|^q$ regularization term in problem (5) is non-lipschitz at 0 when $0 < q < 1$ [20], Lagrange multipliers lose their geometric meaning and therefore KKT optimality condition in standard smooth and non-smooth optimization theories cannot apply to its analysis. But note that the first-order directional derivative of $|x_i|^q$ at 0 still exists if it is allowed to take infinite value, i.e. $\lim_{t\to 0^+} t^{q-1}|d_i|^q = +\infty$. Denote the second term in the objective function of problem (5) as $g(\boldsymbol{v})$. Thus the following analysis will be based on the first-order directional derivatives.

We first give a sufficient optimality condition for problem (5) in the context of ADMM. For simplicity, define $h(\boldsymbol{x}, \boldsymbol{u}) = f(\boldsymbol{x}, \boldsymbol{u}) + g(\boldsymbol{u})$. We say an optimization problem $min\ f(\boldsymbol{x})\ s.t.\ \boldsymbol{x} \in S$ is generic if the first-order necessary optimality condition $f^{'}(\overline{\boldsymbol{x}}; \boldsymbol{d}) \geq 0,\ \forall \boldsymbol{d} \in T_S(\overline{\boldsymbol{x}})$ is also sufficient for $\overline{\boldsymbol{x}}$ to be a local minimum. Requesting a problem to be generic is equivalent to asking that the problem is solvable in the sense of finding a local minimum by first-order optimality condition based algorithms, e.g. IPM which are known robust to deal with OPF non-convexity in practice.

Theorem 1: $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}})$ is a local minimum of problem (5) if given $\rho > 0$, there exists $\overline{\boldsymbol{y}}$, such that ① $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ is a local minimum of $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$; ② $\overline{\boldsymbol{v}}$ is a global minimum of $L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \boldsymbol{v}, \overline{\boldsymbol{y}})$; ③ $\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}} = \boldsymbol{0}$; ④ problem (3) is generic.

Remark 1: Theorem 1 acts as the first-order sufficient optimality condition in the context of ADMM. It allows us to analyze the algorithm without considering the detailed optimality conditions of two sub-problems. The following convergence analysis will be based on this optimality condition.

Then we present a weak result for ADMM-IPM-STO convergence with $0 < q \leq 1$ in the following theorem.

Theorem 2: Let $\{(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k, \boldsymbol{y}^k)\}$ be a sequence generated by ADMM-IPM-STO. Assume that ① problem (3) is generic; ② the sequence $\{\boldsymbol{y}^k\}$ converges to a point, i.e. $\lim_{k\to+\infty} \boldsymbol{y}^k = \overline{\boldsymbol{y}}$. Then $\{(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k)\}$ converge to a limit point $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}})$ which is a local minimum of problem (5).

Remark 2: Theorem 2 actually reveals that if the convergence of the dual variable $\boldsymbol{y}^k$ is observed, it is safe to say that ADMM-IPM-STO achieves a local minimum of problem (5). This theorem substantiates the optimality of the solutions obtained by ADMM-IPM-STO in practice.

Finally, since $g(\boldsymbol{v})$ is convex for $q = 1$, we can obtain stronger convergence result under proper assumptions. We present the strong result in the following theorem.

Theorem 3: Let $\{(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k, \boldsymbol{y}^k)\}$ be a sequence generated by ADMM-IPM-STO. $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$ satisfies the sufficient optimality in Theorem 1. Assume that, for sufficient large k: ① problem $min\ f(\boldsymbol{x}, \boldsymbol{u}) + \left(\boldsymbol{y}^k\right)^T \boldsymbol{u},\ s.t.\ (\boldsymbol{x}, \boldsymbol{u}) \in S$ is generic; ② $(\boldsymbol{x}^k, \boldsymbol{u}^k)$ is in the attraction basin of local minimum $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ of $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$, i.e. $L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}}) \leq L_\rho(\boldsymbol{x}^k, \boldsymbol{u}^k, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$ always holds; ③ $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ is in the attraction basin of local minimum $(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1})$ of $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}^k, \boldsymbol{y}^k)$, i.e. $L_\rho(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}, \boldsymbol{v}^k, \boldsymbol{y}^k) \leq L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \boldsymbol{v}^k, \boldsymbol{y}^k)$ always holds. Then $\{(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k, \boldsymbol{y}^k)\}$ converges to $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$.

Remark 3: Theoretically, the IPM algorithm for OPF problems can only achieve a local minimum. So assumption ② and ③ in theorem 3 actually ensure that all the local minimums are in a single attraction basin and thus the inequality relations do not compromise.

IV. General FACTS Devices Allocation Problems

In our formulation, current mismatch equations are chosen as equality constraints instead of power mismatch equations. Bus voltages and generator current injections are taken as state variables. Consequently, generators and loads are modeled as complex current injections at their buses. All FACTS devices are modeled as parametric complex current injections at related buses. The reasons for these choices are as follows. At first, three series controllable parameters relate to one line in general FACTS devices allocation problems, which leads to a very high-order power balance equation. Subsequently, solving the second-order derivatives is far more difficult in power balance equations than that in current balance equations. Secondly, as every bus or line can be a candidate location for FACTS devices implies a very large number of controllable parameters, and solving the second-order derivatives in conventional formulation thus becomes rather impractical.

A. Branch and FACTS Devices Modeling

Without loss of generality, every bus or line in power systems is considered as a candidate location for FACTS device placement. A general branch model is shown in Fig.1 which is similar to that in [30]. r, x and b are transmission line parameters. $\hat{v}_f$, $\hat{v}_t$, $\hat{i}_f$ and $\hat{i}_t$ are complex voltage and current at "from" and "to" ends of the branch. The series controllable parameters κ, φ and τ are used to describe the effect of TCSC (thyristor control led series capacitor), TCPS (thyristor controlled phase shifter) and ULTC (under load tap changer), respectively. In addition, Shunt compensation devices can be simply modeled as extra susceptance at certain buses which are not shown in Fig. 1. Because this model contains all network parameters, other types of FACTS devices can be equivalently transformed into this model. In particular, the models of STATCOM [31], SSSC [32] and UPFC [33] are special cases of the general model used in this paper.

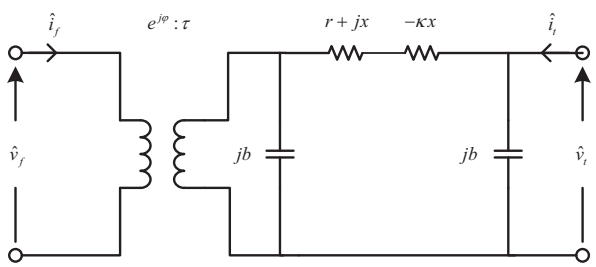


Fig. 1. General Branch Model

From Fig. 1, the relationship between complex voltages and currents at both ends is given by

$$\begin{bmatrix}\hat{i}_f\\ \hat{i}_t\end{bmatrix}=\begin{bmatrix}\tau^2(\hat{y}^*+jb) & -\tau e^{j\varphi}\hat{y}^*\\ -\tau e^{-j\varphi}\hat{y}^* & \hat{y}^*+jb\end{bmatrix}\begin{bmatrix}\hat{v}_f\\ \hat{v}_t\end{bmatrix} \tag{14}$$

where $\hat{y}^*=1/\left(r+j(1-\kappa)x\right)$.

To facilitate optimization computation, the equivalent current injection model is derived and adopted. In this model, the effects of all the controllable parameters are represented by the current injections at the "from" and "to" ends of the branch and this model possesses the versatility to adapt to various types of FACTS devices. Assume that Fig. 1 is equivalent to Fig. 2 in which equivalent complex current injections $\Delta\hat{i}_f$ and $\Delta\hat{i}_t$ are introduced to depict the effects of all FACTS devices and $\hat{y}=1/\left(r+jx\right)$. The following equation holds:

$$\begin{bmatrix}\hat{i}_f+\Delta\hat{i}_f\\ \hat{i}_t+\Delta\hat{i}_t\end{bmatrix}=\begin{bmatrix}\hat{y}+jb & -\hat{y}\\ -\hat{y} & \hat{y}+jb\end{bmatrix}\begin{bmatrix}\hat{v}_f\\ \hat{v}_t\end{bmatrix} \tag{15}$$

Combining (14) with (15), we can acquire parametric complex

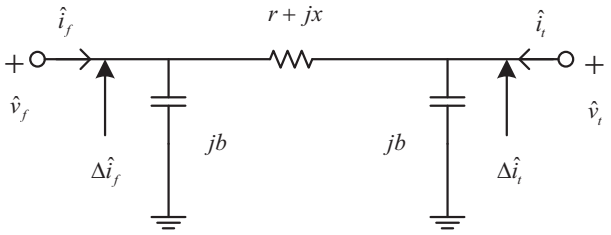


Fig. 2. Equivalent Current Injection Model

current injections as

$$\begin{bmatrix}\Delta\hat{i}_f\\ \Delta\hat{i}_t\end{bmatrix}=\begin{bmatrix}\hat{y}-\tau^2\hat{y}^*+jb(1-\tau^2) & -\hat{y}+\tau e^{j\varphi}\hat{y}^*\\ -\hat{y}+\tau e^{-j\varphi}\hat{y}^* & \hat{y}-\hat{y}^*\end{bmatrix}\begin{bmatrix}\hat{v}_f\\ \hat{v}_t\end{bmatrix} \tag{16}$$

Equation (16) shows that $\Delta\hat{i}_f$ and $\Delta\hat{i}_t$ are the functions of series controllable parameters and complex voltages at both ends of the line and they can completely represent the effect of controllable devices.

B. Sparsity-constrained OPF Model

In the sparsity-constrained OPF problems, decision variables are divided into two groups and denoted by two vectors. Decision vector $\boldsymbol{u}$ consists of setting values of all the candidate FACTS devices, i.e.

$$\boldsymbol{u}=(\boldsymbol{b}_{sh},\ \boldsymbol{\varphi},\ \boldsymbol{\kappa}) \tag{17}$$

where $\boldsymbol{b}_{sh}\in\mathbf{R}^{n_b}$ denotes the susceptance of shunt compensation devices at every bus; $\boldsymbol{\varphi}\in\mathbf{R}^{n_l}$ and $\boldsymbol{\kappa}\in\mathbf{R}^{n_l}$ denote the shift angle and series compensation rate at every line, respectively. Certain component of $\boldsymbol{u}$ being zero indicates no corresponding device is installed. Vector $\boldsymbol{x}$ contains bus voltages, generator current injections and UTLC ratios, i.e.

$$\boldsymbol{x}=(\boldsymbol{e},\ \boldsymbol{f},\ \boldsymbol{a}_g,\ \boldsymbol{b}_g,\ \boldsymbol{\tau}) \tag{18}$$

where $\boldsymbol{e}\in\mathbf{R}^{n_b}$ and $\boldsymbol{f}\in\mathbf{R}^{n_b}$ denote the real and imaginary parts of bus voltages; $\boldsymbol{a}_g\in\mathbf{R}^{n_g}$ and $\boldsymbol{b}_g\in\mathbf{R}^{n_g}$ denote the real and imaginary parts of generator current injections; $\boldsymbol{\tau}\in\mathbf{R}^{n_l}$ denotes the transformer ratio at every line. Between them, vector $\boldsymbol{u}$ is expected to be very sparse after optimization due to economic consideration.

The equality constraints of OPF formulation in this paper are the network current mismatch equations. Let $\hat{\boldsymbol{h}}\in\mathbf{C}^{n_b}$ denotes the complex current mismatch at every bus. Its i^{th} elment is given by

$$\hat{h}_i(\boldsymbol{x},\boldsymbol{u})=\sum_{k\in N_{gi}}(a_{gk}+jb_{gk})+\left(\frac{p_{li}+jq_{li}}{e_i+jf_i}\right)^c+\Delta\hat{i}_i-\sum_{k=1}^{n_b}\hat{y}_{ik}(e_k+jf_k) \tag{19}$$

where N_{gi} is the index set of generators installed at bus i; p_{li} and q_{li} denote the active and reactive power of load at bus i, respectively; $\hat{y}_{ik}$ is the $(i,k)^{\text{th}}$ element of the nodal admittance matrix of the original network shown in Fig. 2. $\Delta\hat{i}_i$ is the complex current injection induced by all the controllable equipments related to bus i. Define N_{fi} (N_{ti}) as the index set of branches which takes bus i as their "from" ("to") end. Then $\Delta\hat{i}_i$ is expressed as

$$\Delta\hat{i}_i=\sum_{k\in N_{fi}}\Delta\hat{i}_{fk}+\sum_{k\in N_{ti}}\Delta\hat{i}_{tk}-jb_{shi}(e_i+jf_i) \tag{20}$$

where $\Delta\hat{i}_{fk}$ $(\Delta\hat{i}_{tk})$ is the complex current injection induced by related controllable devices at the "from" ("to") end of the branch k which is given by (16). The last term in (20) denotes the complex current injection induced by shunt compensation devices at bus i. Therefore, the equality constraints can be written in a compact form as

$$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{u})=\left(\text{Re}\left(\hat{\boldsymbol{h}}(\boldsymbol{x},\boldsymbol{u})\right),\ \text{Im}\left(\hat{\boldsymbol{h}}(\boldsymbol{x},\boldsymbol{u})\right)\right)=\boldsymbol{0} \tag{21}$$

The inequality constraints are steady state security constraints, including current magnitude limit for every line and voltage magnitude limit for every bus

$$\|(\hat{y}+jb)(e_f+jf_f)-\hat{y}(e_t+jf_t)\|_2^2\le(i_{max})^2 \tag{22}$$

$$(v_{mim})^2\le\|e_i+jf_i\|_2^2\le(v_{max})^2 \tag{23}$$

and physical limits of devices, including generator active and reactive power output:

$$p_{kmim}\le\text{Re}\left((e_i+jf_i)(a_{gk}+jb_{gk})^c\right)\le p_{kmax} \tag{24}$$

$$q_{kmim}\le\text{Im}\left((e_i+jf_i)(a_{gk}+jb_{gk})^c\right)\le q_{kmax} \tag{25}$$

where generator k is installed at bus i; setting value limits of ULTC, SVC, TCPS and TCSC:

$$\tau_{kmim}\le\tau_k\le\tau_{kmax} \tag{26}$$

$$b_{shkmim}\le b_{shk}\le b_{shkmax} \tag{27}$$

$$\varphi_{kmim}\le\varphi_k\le\varphi_{kmax} \tag{28}$$

$$\kappa_{kmim}\le\kappa_k\le\kappa_{kmax} \tag{29}$$

Thus inequality constraints are written in a compact form as

$$\boldsymbol{g}_{mim}\le\boldsymbol{g}(\boldsymbol{x},\boldsymbol{u})\le\boldsymbol{g}_{max} \tag{30}$$

In the proposed formulation, it is certainly free to choose various kinds of objective functions. Without loss of generality, system loadability is taken as objective function in our case studies to test the proposed algorithm. The complex power injection of loads at bus i is modified as

$$p_{li} + jq_{li} = \eta(p_{li0} + jq_{li0}) \tag{31}$$

where η is the loadability factor and $p_{li0} + jq_{li0}$ is the initial complex power injection of loads at bus i. Hence, the objective function is given by

$$f(\boldsymbol{x}, \boldsymbol{u}) = -\eta \tag{32}$$

To sum up, the sparsity-constrained OPF formulation for general FACTS devices allocation problems can be written in a compact form as (3) where feasible set S are specified by (21) and (30). Objective function is defined in (32). Problem (3) can be solved with the ADMM-IPM-STO algorithm discussed in Section II.

V. Case Studies

To validate the algorithm proposed, single-type and multiple-type FACTS devices allocation problems have been tested on standard IEEE 30-buses, 118-buses and 300-buses system, respectively. The system data is extracted from Matpower 4.1. ADMM-IPM-STO algorithm was programmed in MATLAB running on a Win7 PC with Intel Core i5 1.80-GHz CPU and 4 GB of RAM.

A. Parameter Settings

The value of q of L_q norms is an economical parameter related to the investment cost of different types of FACTS device. For practical application, L_q norms can choose three values, $L_{1/2}$, $L_{2/3}$ and L_1. If the investment cost of a certain type of FACTS devices is dominated by the number of devices, $L_{1/2}$ norm is a better choice. If the setting value of the FACTS devices play a major role in the investment cost, L_1 norm is preferred. $L_{2/3}$ norm acts as a compromise between $L_{1/2}$ norm and L_1 norm. In our experiment, $L_{1/2}$ norm is used in multiple-type FACTS devices allocation and single-type TCPS allocation; L_1 norm is applied to single-type SVC allocation and TCSC allocation.

The physical meaning of regularization parameter λ represents the cost of per unit FACTS devices measured in L_q norm and its value should be decided by the decision maker based on their specification. To test the robustness of the algorithm proposed, the sparsity-constrained OPF problems are solved with decreasing values of λ, which means system loadability is increased through installation of an increasing number of FACTS devices. In other words, every λ value is associated with a devices allocation strategy. Due to space limitation, only under a certain device number, the allocation strategy with the largest loadability factor is given to illustrate the relationship among device numbers, total installed capacity and system loadability.

Theoretically, the value of the augmented Lagrangian parameter ρ will not affect the result of the algorithm only if it surpasses a threshold value which is problem-dependent and unknown before the problem solved. Nevertheless, the value of ρ affects the convergence process of the algorithm. Large values of ρ place a severe penalty on violations of primal feasibility and thus tend to produce small primal residuals. Conversely, small values of ρ tend to reduce the dual residual at the cost of the primal residual. Therefore, an ideal value of ρ should keep the primal and the dual residuals within moderate difference as they both converge to zero. ρ is usually chosen by cross-validation.

In addition, the decision variable $\boldsymbol{u}$ is usually rescaled by multiplying its components with some factors. For example, considering different line distances, every κ is multiplied by the reactance of the related line. In multi-type FACTS devices allocation problems, to reflect the price differences among different types of FACTS devices, b_{sh}, κ and φ are multiplied by different factors α_c, α_κ and α_φ, respectively. In our study, the allowable range of the compensation rate of TCSCs, the shift angle of TCPSs and the ratio of UTLCs are $0\% \sim 50\%$, $-15^\circ \sim +15^\circ$ and $0.9 \sim 1.1$, respectively. The setting values of SVC are unbounded. ϵ_{primal} and ϵ_{dual} are both set to be 10^{-4}.

B. Multiple-type FACTS Devices Allocation

Multiple-type FACTS devices allocation is conducted on IEEE 30-bus system. The candidate FACTS devices are SVC, TCSC and TCPS. In this problem, $\boldsymbol{x}$ is 72 dimensional and $\boldsymbol{u}$ is 112 dimensional. By running a conventional OPF without FACTS devices, the maximum loadability of IEEE 30-bus system is 1.020. Regardless of device costs, i.e. setting $\lambda = 0$ in our algorithm, the theoretical maximum loadability by FACTS device installation is 1.735, and the results under this condition are taken as the initial values of decision variables in problems with other values of λ. Thus, during the optimization process, the loadability factor η, the device number, setting values decrease from the original values to achieve primal and dual convergence of the ADMM-IPM-STO algorithm. The convergence processes of primal and dual errors are shown in Fig. 3, and the changing process of the loadability factor is shown in Fig. 4 with $\lambda = 0.29$, $\rho = 500$, $\alpha_c = 0.1$, $\alpha_\kappa = 20$ and $\alpha_\varphi = 200$.

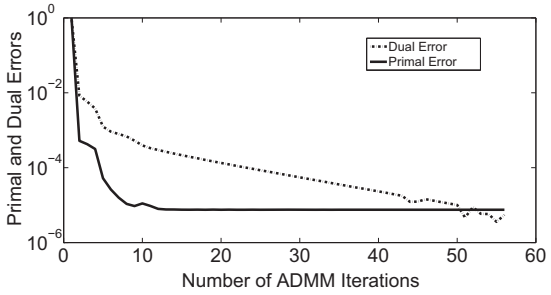


Fig. 3. Primal and Dual Residuals Convergence Process

Fig.3 shows that the primal residual stops decreasing after about 10 iterations, as the NLP sub-problem precision restriction is set to be 10^{-5}. The dual convergence is much slower than the primal convergence. Essentially, the reason for this phenomenon is that, in the method of multipliers, the primal

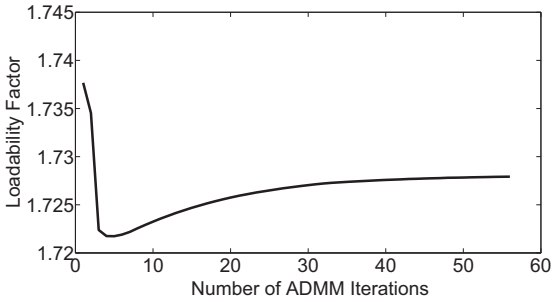


Fig. 4. Loadability Factor Changing Process

TABLE I
MULTIPLE-TYPE FACTS DEVICES ALLOCATION STRATEGY ON 30-BUS SYSTEM

λ	No.	η	Allocation Strategy
3.38	1	1.539	TCSC:10(50%)[a]
2.67	2	1.608	SVC:8(90.5)[b]; TCSC:10(50%)
0.30	3	1.723	SVC:8(33.1); TCSC:10(36.0%),29(49.5%)
0.29	4	1.728	SVC:8(46.0),28(25.2); TCSC:10(34.1%),29(50%)

[a]Line number (compensation rate)
[b]Bus number (Var compensation capacity in MVar)

problem is solved with a second-order method (IPM) whereas the dual problem with a first-order method (steepest ascent) [34]. Fig. 4 shows the loadability factor declines from the initial value and finally reaches the steady value 1.728 as convergence achieved. After optimization, only two SVCs and two TCSCs, among 112 candidates, are selected and installed in the network, shown in the last row of Table I.

Furthermore, the allocation strategies with device number from 1 to 4 are listed in Table I. It is obvious that, with only 4 FACTS devices installed, the loadability factor can reach 1.728, up to 99.6% of the theoretical maximum. This validates the sparse feature of FACTS devices allocation problem. In other words, a large number of FACTS devices are not only uneconomical but also unnecessary.

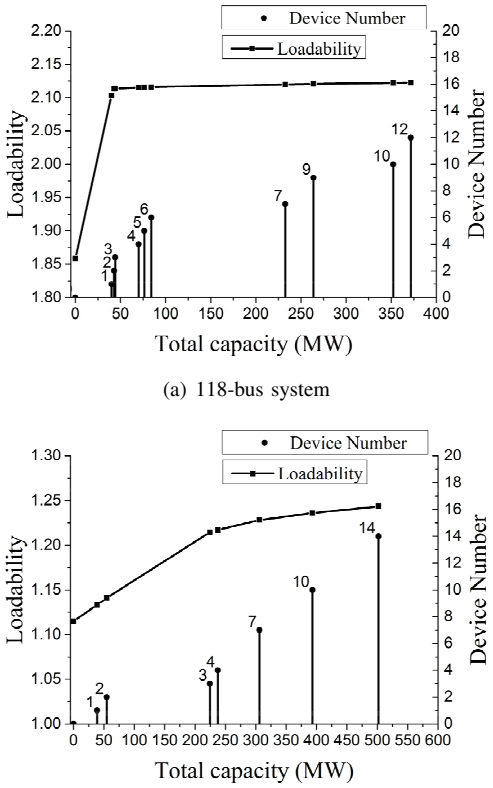
C. Single-Type FACTS Devices Allocation

SVC, TCSC and TCPS allocation problems are conducted on IEEE 118- and 300-bus systems. The basic information of all the six problems is summarized in Table II.

Results of all the six problems are shown in Fig.5 to Fig. 7, which illustrate the evolution of system loadability as the device number and total installed capacity increase. In all three types of FACTS allocation problems studied in this paper, the maximum loadability saturates after several devices installed (typical number of devices is 3), and the marginal utility of FACTS devices significantly diminishes.

TABLE II
SINGLE-TYPE FACTS DEVICES ALLOCATION PROBLEMS

Problems	Dim. $\boldsymbol{x}$	Dim. $\boldsymbol{u}$	η_{max} no device	theoretical η_{max}
SVC118	353	118	1.859	2.124
TCSC118	353	186	1.859	2.283
TCPS118	353	186	1.859	1.867
SVC300	845	300	1.116	1.243
TCSC300	845	411	1.116	1.236
TCPS300	845	411	1.116	1.120



(a) 118-bus system

(b) 300-bus system

Fig. 5. Loadability enhancement by optimal allocation of SVCs on 118 and 300 bus systems

TABLE III
SVC ALLOCATION STRATEGY ON 118-BUS SYSTEM

No.	η	Allocation Strategy	η/η_{max}
1	2.103	76(40.12)[a]	99.0%
2	2.113	76(33.96), 118(8.89)	99.5%
3	2.114	76(34.06), 95(1.28), 118(8.82)	99.5%

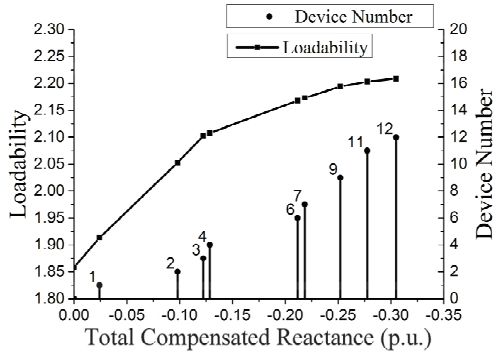
[a]Bus number (Var compensation capacity in MVar)

Detailed allocation strategies for SVC allocation on IEEE 118-bus system, TCSC allocation on IEEE 300-bus system and TCPS allocation on IEEE 118-bus system are listed in Table III, Table IV and Table V, respectively. Only the parameters and locations for the first 3 devices are given, as it is clearly shown in Fig.5 to Fig. 7 that the enhancement of loadability will saturate with more than 3 devices. This again demonstrates the sparsity of FACTS devices allocation problems.

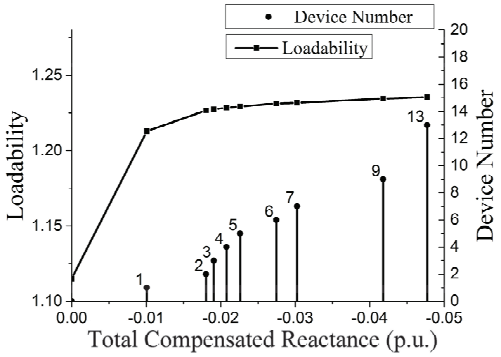
TABLE IV
TCSC ALLOCATION STRATEGY ON 300-BUS SYSTEM

No.	η	Allocation Strategy	η/η_{max}
1	1.213	177(37%)[a]	98.1%
2	1.226	177(50%), 367(26%)	99.2%
3	1.227	1(12%), 177(50%), 367(31%)	99.3%

[a]Line number (compensation rate)

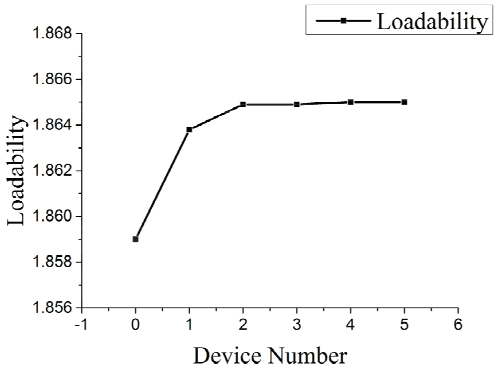


(a) 118-bus system

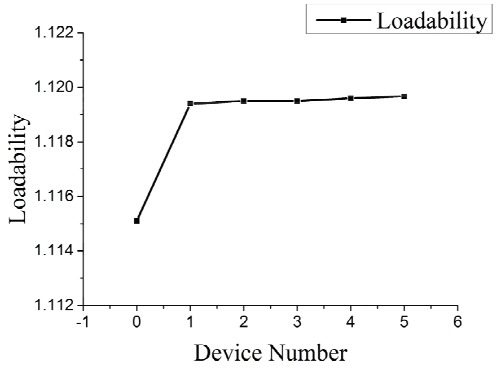


(b) 300-bus system

Fig. 6. Loadability enhancement by optimal allocation of TCSCs on 118 and 300 bus systems



(a) 118-bus system



(b) 300-bus system

Fig. 7. Loadability enhancement by optimal allocation of TCPSs on 118 and 300 bus systems

TABLE V
TCPS ALLOCATION STRATEGY ON 118-BUS SYSTEM

No.	η	Allocation Strategy	η/η_{max}
1	1.864	123(-5.418)[a]	99.8%
2	1.865	121(3.448), 123(-4.189)	99.9%
3	1.865	121(1.932), 122(1.590), 123(-4.113)	99.9%

[a]Line number (shift angle in degree)

VI. RESULT DISCUSSION

A. Comparisons with Other Methods

Full AC power flow model has been used in our problem formulation. This certainly improves the quality of the solution compared to other methods employing simplified power flow models. The DC flow used in MILP method [12] neglects the nonlinearity of the power flow and constraints on voltage magnitude and cannot precisely reflect the limits on line capacity. The LFB flow used in MIQP method [2] has also been simplified by transforming the quadratic equality constraints into linear inequality constraints. Therefore the nonlinearity of power flow is not completely represented. The modeling of limits on line capacity is also deficient. The GA method in [8] uses a simplified version of AC flow by neglecting transverse conductance of transmission lines. The motivation for all those simplifications is either convenience of adopting certain methods or simplifying numerical computation. But this will inevitably affect the quality of the solutions. For example, we have observed in our experiment, the constraints of voltage magnitude and line capacity often act as binding constraints at the final solution which indicates that defects in representing these constraints surely change the optimal solution. These simplified methods are, to some extent, eligible for preliminary planning, but their results need to be verified by the full AC model. Although the different problem formulations have complicated influence on the final solutions, we would like to make some rough comparison as follows. For SVC allocation on 300-bus system, we install 3 SVCs to improve the loadability to 1.217 while [6] improves the loadability to 1.207 with 5 SVCs. For TCSC allocation on 300-bus system, the loadability is improved to 1.227 with 3 TCSCs in this paper compared to 1.081 with 19 TCSCs in [2]. For TCPS allocation on 118-bus system, the loadability achieves 1.865 with 3 TCPSs by our approach compared to 1.76 with 13 TCPSs in [12]. These comparisons show that our approach is generally more effective to identify the optimal locations and setting values of FACTS devices.

To roughly evaluate the efficiency of the proposed algorithm, we continue the comparison originally conducted in [2] and list the results in Table VI. At first, the proposed ADMM-IPM-STO employs the full AC power flow model rather than simplified models such as the DC model and the LFB model used in MILP [12] and MIQP [2], which improves the accuracy and reliability of the computation results. Secondly, the ADMM-IPM-STO is far more computationally efficient than GA [8] and also offers faster or at least comparable performance compared with MIQP [2].

TABLE VI
ALGORITHM EFFICIENCY COMPARISON

Methods	Network	Device	Model	Machine	Time
GA [8]	200-bus	TCPS	AC	Sun SPARC Workstation	1.5h
MILP [12]	300-bus	TCPS	DC	PC 450MHz 128MB RAM	2.5s
MIQP [2]	300-bus	TCSC	LFB	Dell OptiPlex GX520	218s
Proposed	300-bus	TCPS	AC	PC 1.80GHz 3.85GB RAM	63s

The major competitors of the proposed method are those mixed-integer programming based methods, including MILP [12], MIQP [2] and MINP [6]. In those methods, branch-and-cut or benders decomposition are involved to tackle the binary variables and form a series of continuous subproblems. The continuous subproblems can be linear programming, quadratic programming or nonlinear programming according to their problem formulations. In those methods, the number of continuous subproblems is strongly related to the number of binary variables. Thus the computation time will significantly increase as the allowable device number or the system scale increase. We compare the computation time of the proposed method with MINP and MIQP in Fig.8 and Fig.9 as allowable device number and system scale increase, respectively. These two graphs show that the computation time of mixed-integer programming based methods are very sensitive to allowable device number and system scale. Whereas the computation time of ADMM-IPM-STO is almost irrelevant to allowable device number and far less sensitive to problem scale.

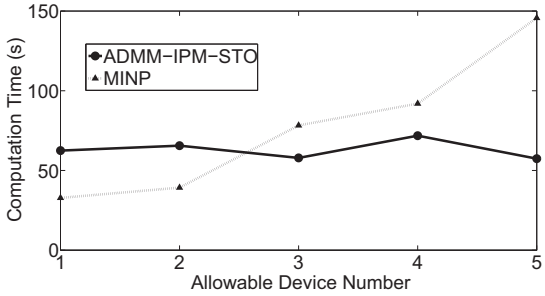


Fig. 8. Computation time comparison between ADMM-IPM-STO and MINP as allowable device number increase for SVC allocation on 300-bus system

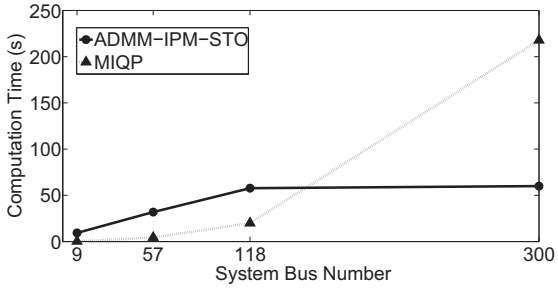


Fig. 9. Computation time comparison between ADMM-IPM-STO and MIQP as system scale increase for TCSC allocation problem

B. Convergence and Non-convexity

Although so far general convergence result of ADMM-IPM-STO for all $q \in (0, 1]$ has not been established, our analysis shows the optimality of solutions obtained in practice is theoretically substantiated, and the convergence for $q = 1$ is quite well guaranteed. Furthermore, the convergence of proposed algorithm is very robust in practice for each value of q with closed-form solutions, including $q = 1/2$, $q = 2/3$ and $q = 1$. In our experiment, ADMM-IPM-STO can achieve convergence for every type of FACTS allocation problems as long as the parameter ρ is adjusted to a proper range. The rule of thumb of finding the proper ρ, in our experience, is to make the primal convergence a little faster than dual convergence. Actually, by only several trial-and-error processes a suitable parameter can then be found because the range of applicable ρ is usually very wide and the dependency of the algorithm performance on the value of ρ is pretty loose.

Theoretically, ADMM-IPM-STO can only achieve local minimum due to the non-convex problem formulation. Note that there are two sources of non-convexity: the original non-convexity of OPF problems and the non-convexity of L_q norms. For the non-convexity of OPF problems, because we have applied IPM to solve the OPF subproblems, ADMM-IPM-STO shares the same limitation of the IPM, and therefore only local minimum can be expected. Recently, there have been some attempts to obtain global solution of OPF problems. The author of [35] established a sufficient condition for zero duality gap of the semi-definite programming formulation of the dual problem of OPF, which leads to global solution to OPF problems. But this sufficient condition does not always hold in every network [36]. None of the current methods can guarantee the global optimum solution of general OPF problems. The proposed ADMM-IPM-STO thus is also incapable of guaranteeing global solution. For the additional non-convexity of L_q norms, a natural question is that whether this additional non-convexity will produce additional local optimums. We have conducted a series of numerical experiments to answer this question. For all tests, we fix the initial value of $\boldsymbol{x}$ and randomly select different initial value of $\boldsymbol{u}$. This means the initial condition of the original network is fixed while the initial guess of the FACTS device allocation strategy changes. Test results show that all different initial values of $\boldsymbol{u}$ result in a common optimal solution. This shows that the non-convexity brought by L_q does not produce additional local minimum. The fundamental reason for this is that the STOs give global optimal solutions to L_q regularization problems.

VII. CONCLUSION

This paper has proposed a novel formulation and algorithm for FACTS devices allocation problems. Based on the sparse characteristics of device placement, FACTS allocation problems have been formulated as a sparsity-constrained OPF problem. An ADMM-IPM-STO algorithm, which combines the state-of-art algorithms in both sparse optimization and OPF, has been proposed to simultaneously determine the numbers, locations, setting values and types of FACTS devices. $L_q(0 < q < 1)$ norms have been firstly introduced to represent

sparsity in FACTS devices allocation problems, in which q is an economical parameter related to the investment cost of different types of FACTS device. The proposed method has been tested on several IEEE standard systems in both multiple-type allocation problem as well as single-type problems, respectively. Case studies demonstrate the effectiveness and efficiency of this method.

In addition, the proposed formulation and algorithm can be easily applied to other objective functions since we have presented a general sparsity-constrained OPF model and the same calculating process of the ADMM-IPM-STO can be used no matter what objective function is chosen. Moreover, the proposed approach can be extended to FACTS allocation problems considering multiple contingencies by further incorporating the ideas of "security-constrained OPF" [29] and "group sparsity" [37]. FACTS devices allocation problem considering multiple contingencies can be formulated as

$$\min_{(\boldsymbol{x}_c,\boldsymbol{u}_c):\forall c\in\mathcal{C}} f(\boldsymbol{x}_0,\boldsymbol{u}_0)+\lambda\sum_{k=1}^{n}\|\boldsymbol{w}_k\|_p^q \tag{33a}$$

$$\boldsymbol{h}_c(\boldsymbol{x}_c,\boldsymbol{u}_c)=\boldsymbol{0} \qquad \forall c\in\mathcal{C} \tag{33b}$$

$$\boldsymbol{g}_{mim}\le \boldsymbol{g}_c(\boldsymbol{x}_c,\boldsymbol{u}_c)\le \boldsymbol{g}_{max} \qquad \forall c\in\mathcal{C} \tag{33c}$$

where $\mathcal{C}=\{0,1,2,\dots,t\}$ is the set of prespecified $t+1$ contingencies with $c=0$ representing the base case; for each contingency, $\boldsymbol{u}_c$, $\boldsymbol{x}_c$, $\boldsymbol{h}_c(\cdot)$ and $\boldsymbol{g}_c(\cdot)$ share the same definition as (17), (18), (21) and (30), respectively; $\boldsymbol{w}_k=(u_{0k},u_{1k},\dots,u_{tk})$ are the setting values of the k-th device in all contingencies. The regularization term in the objective function is referred as $L_{q,p}$ norm with $q\in(0,1]$ and $p\in[2,+\infty]$. This regularizer leads the setting values of the same FACTS device in different contingencies to be zeros or non-zeros simultaneously which enforces sparsity on installation sites.

Appendix A
Shrinkage-thresholding Operators

The specific expression of STOs for $L_{1/2}$ [20], $L_{2/3}$ [19] and L_1 [21] norms are presented here. They are all diagonally nonlinear operators uniformly expressed as

$$\text{Shrink}(\boldsymbol{z},\beta)=[H_\beta(z_1)\quad H_\beta(z_2)\dots\quad H_\beta(z_n)]^T \tag{34}$$

For $q=1/2$,

$$H_\beta(z_i)=\begin{cases}\frac{2}{3}|z_i|\left(1+\cos\left(\frac{2\pi}{3}-\frac{2\varphi_\beta(z_i)}{3}\right)\right) & z_i>p(\lambda)\\ 0 & |z_i|\le p(\lambda)\\ -\frac{2}{3}|z_i|\left(1+\cos\left(\frac{2\pi}{3}-\frac{2\varphi_\beta(z_i)}{3}\right)\right) & z_i<-p(\lambda)\end{cases} \tag{35}$$

where $\varphi_\beta(z_i)=\arccos\left(\frac{\beta}{8}\left(\frac{|z_i|}{3}\right)^{-\frac{2}{3}}\right)$ and $p(\lambda)=\frac{\sqrt[3]{54}}{4}(\beta)^{\frac{2}{3}}$.

For $q=2/3$,

$$H_\beta(z_i)=\begin{cases}\left(\frac{m+\sqrt{2|z_i|/m-m^2}}{2}\right)^3 & z_i>p(\beta)\\ 0 & |z_i|\le p(\beta)\\ -\left(\frac{m+\sqrt{2|z_i|/m-m^2}}{2}\right)^3 & z_i<-p(\beta)\end{cases} \tag{36}$$

where $m=\frac{2}{\sqrt{3}}\beta^{\frac{1}{4}}\left(\cosh\left(\frac{\phi}{3}\right)\right)^{\frac{1}{2}}$, $\phi=\text{arccosh}\left(\frac{27z_i^2}{16}\beta^{-\frac{3}{2}}\right)$ and $p(\beta)=\frac{2}{3}(3\beta^3)^{\frac{1}{4}}$.

For $q=1$,

$$H_\beta(z_i)=\text{sgn}(z_i)\max\{|z_i|-\beta,0\} \tag{37}$$

Appendix B
A Lemma

We present a Lemma which will be frequently used in Appendix C, D and E.

Lemma 1: Assume that $\overline{\boldsymbol{x}}$ is a local minimum of $min\ h_1(\boldsymbol{x})+h_2(\boldsymbol{x}), s.t.\ \boldsymbol{x}\in S$; the first-order directional derivative $h_1'(\overline{\boldsymbol{x}};\boldsymbol{d})$ exists; $h_2(\boldsymbol{x})$ is continuously differentiable and $\nabla^2 h_2(\overline{\boldsymbol{x}})=\boldsymbol{0}$; If problem $min\ h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x}\in S$ is generic, then $\overline{\boldsymbol{x}}$ is also a local minimum of $h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x}\in S$.

Proof: Since $\overline{\boldsymbol{x}}$ is a local minimum of $h_1(\boldsymbol{x})+h_2(\boldsymbol{x}),\ s.t.\ \boldsymbol{x}\in S$, we have $h_1'(\overline{\boldsymbol{x}};\boldsymbol{d})+\nabla h_2(\overline{\boldsymbol{x}})^T d\ge\boldsymbol{0},\forall d\in T_S(\overline{\boldsymbol{x}})$ where $T_S(\overline{\boldsymbol{x}})$ denotes the tangent cone of S at $\overline{\boldsymbol{x}}$. It follows from $\nabla h_2(\overline{\boldsymbol{x}})=\boldsymbol{0}$ that $h_1'(\overline{\boldsymbol{x}};\boldsymbol{d})\ge\boldsymbol{0},\forall d\in T_S(\overline{\boldsymbol{x}})$. Because problem $min\ h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x}\in S$ is generic, $\overline{\boldsymbol{x}}$ is also a local minimum of $h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x}\in S$. ■

Appendix C
Proof of Theorem 1

Proof: Under condition ① and ②, considering condition ③, there exists $\delta>0$, $\forall(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in F\cap B\left((\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}}),\delta\right)$ such that $L_\rho(\boldsymbol{x},\boldsymbol{u},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})\ge f(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})$ and $L_\rho(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\boldsymbol{v},\overline{\boldsymbol{y}})\ge g(\overline{\boldsymbol{v}})$. Adding the two inequalities and again noticing condition ③, we have $h(\boldsymbol{x},\boldsymbol{u})+\rho\|\boldsymbol{u}-\overline{\boldsymbol{u}}\|_2^2\ge h(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})$, i.e. $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})$ is a local minimum of $h(\boldsymbol{x},\boldsymbol{u})+\rho\|\boldsymbol{u}-\overline{\boldsymbol{u}}\|_2^2$. Considering assumption ④ and using Lemma 1, $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})$ is also a local minimum of $h(\boldsymbol{x},\boldsymbol{u})$. ■

Appendix D
Proof of Theorem 2

Proof: Note that $(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1})$ is a strictly local minimum of $L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}^k,\boldsymbol{y}^k)$ due to the existence of the quadratic term and assumption ①. Therefore there exists $\alpha>0$ such that $L_\rho(\boldsymbol{x}^k,\boldsymbol{u}^k,\boldsymbol{v}^k,\boldsymbol{y}^k)-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^k,\boldsymbol{y}^k)\ge\alpha\left(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|_2^2+\|\boldsymbol{u}^k-\boldsymbol{u}^{k+1}\|_2^2\right)$. Since $\boldsymbol{v}^{k+1}$ is the global minimum of $L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v},\boldsymbol{y}^k)$, in the same way, there exists $\beta>0$ such that $L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^k,\boldsymbol{y}^k)-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^k)\ge\beta\|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_2^2$. Considering above two inequalities and noticing (9) and denoting $c=\min\{\alpha,\beta\}$, we have

$$\begin{aligned}
&L_\rho(\boldsymbol{x}^k,\boldsymbol{u}^k,\boldsymbol{v}^k,\boldsymbol{y}^k)-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^{k+1})\\
=&L_\rho(\boldsymbol{x}^k,\boldsymbol{u}^k,\boldsymbol{v}^k,\boldsymbol{y}^k)-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^k,\boldsymbol{y}^k)+\\
&L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^k,\boldsymbol{y}^k)-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^k)+\\
&L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^k)-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^{k+1})\\
\ge&c\left(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|_2^2+\|\boldsymbol{u}^k-\boldsymbol{u}^{k+1}\|_2^2+\|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_2^2\right)\\
&-(1/\rho)\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|_2^2
\end{aligned} \tag{38}$$

Taking summation of above inequalities and noticing $L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y})$ is bounded below, it gives

$$c\sum_{k=0}^{+\infty}\left(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|_2^2+\|\boldsymbol{u}^k-\boldsymbol{u}^{k+1}\|_2^2+\|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_2^2\right) - (1/\rho)\sum_{k=0}^{+\infty}\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|_2^2<+\infty \tag{39}$$

According to assumption ②, we have $\sum_{k=0}^{+\infty}\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|_2^2<+\infty$. Therefore

$$\sum_{k=0}^{+\infty}\left(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|_2^2+\|\boldsymbol{u}^k-\boldsymbol{u}^{k+1}\|_2^2+\|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_2^2\right)<+\infty \tag{40}$$

which implies the convergence of $\{(\boldsymbol{x}^k,\boldsymbol{u}^k,\boldsymbol{v}^k)\}$. We then denote the limit point as $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}})$. In other words, $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$ is a stationary point of ADMM-IPM-STO. Therefore, condition ①, ② and ③ in Theorem 1 are satisfied. Furthermore, assumption ① directly leads to condition ④ in Theorem 1. So $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}})$ is a local minimum of problem (5). ■

APPENDIX E
PROOF OF THEOREM 3

Proof: Since $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})$ is a local minimum of $L_\rho(\boldsymbol{x},\boldsymbol{u},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$ and $(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1})$ is in its attraction basin by assumption ②, we have $L_\rho(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})\le L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$. Considering assumption ① and using Lemma 1, we obtain

$$f(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})+\overline{\boldsymbol{y}}^T\overline{\boldsymbol{u}}\le f(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1})+\overline{\boldsymbol{y}}^T\boldsymbol{u}^{k+1} \tag{41}$$

Similarly, $\overline{\boldsymbol{v}}$ is the global minimum of $L_\rho(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\boldsymbol{v},\overline{\boldsymbol{y}})$, so we have $L_\rho(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})\le L_\rho(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\boldsymbol{v}^{k+1},\overline{\boldsymbol{y}})$. Considering the convexity of $g(\boldsymbol{v})$ and using Lemma 1, we obtain

$$g(\overline{\boldsymbol{v}})-\overline{\boldsymbol{y}}^T\overline{\boldsymbol{v}}\le g(\boldsymbol{v}^{k+1})-\overline{\boldsymbol{y}}^T\boldsymbol{v}^{k+1} \tag{42}$$

Add (41) and (42) and notice $\overline{\boldsymbol{u}}-\overline{\boldsymbol{v}}=\boldsymbol{0}$ to obtain

$$f(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})+g(\overline{\boldsymbol{v}})\le f(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1})+g(\boldsymbol{v}^{k+1})+\overline{\boldsymbol{y}}^T\left(\boldsymbol{u}^{k+1}-\boldsymbol{v}^{k+1}\right) \tag{43}$$

According to ADMM-IPM-STO, $(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1})$ is the local minimum of $L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}^k,\boldsymbol{y}^k)$. Considering assumption ③, substituting $\boldsymbol{y}^k=\boldsymbol{y}^{k+1}-\rho(\boldsymbol{u}^{k+1}-\boldsymbol{v}^{k+1})$ and again using Lemma 1, we obtain

$$f(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1})+\left(\boldsymbol{y}^{k+1}+\rho(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k)\right)^T\boldsymbol{u}^{k+1} \le f(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})+\left(\boldsymbol{y}^{k+1}+\rho(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k)\right)^T\overline{\boldsymbol{u}} \tag{44}$$

Similarly, noticing $\boldsymbol{v}^{k+1}$ is the global minimum of $L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v},\boldsymbol{y}^k)$, substituting $\boldsymbol{y}^k=\boldsymbol{y}^{k+1}-\rho(\boldsymbol{u}^{k+1}-\boldsymbol{v}^{k+1})$ and using Lemma 1 yields

$$g(\boldsymbol{v}^{k+1})-\left(\boldsymbol{y}^{k+1}\right)^T\boldsymbol{v}^{k+1}\le g(\overline{\boldsymbol{v}})-\left(\boldsymbol{y}^{k+1}\right)^T\overline{\boldsymbol{v}} \tag{45}$$

In the same way, we also have

$$g(\boldsymbol{v}^{k+1})-\left(\boldsymbol{y}^{k+1}\right)^T\boldsymbol{v}^{k+1}\le g(\boldsymbol{v}^k)-\left(\boldsymbol{y}^{k+1}\right)^T\boldsymbol{v}^k \tag{46}$$

$$g(\boldsymbol{v}^k)-\left(\boldsymbol{y}^k\right)^T\boldsymbol{v}^k\le g(\boldsymbol{v}^{k+1})-\left(\boldsymbol{y}^k\right)^T\boldsymbol{v}^{k+1} \tag{47}$$

We add (46) and (47) to obtain

$$(\boldsymbol{y}^{k+1}-\boldsymbol{y}^k)^T(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k)\ge 0 \tag{48}$$

Adding (43), (44) and (45), rearranging (see Appendix F), and noticing (48), we can obtain

$$w^k-w^{k+1}\ge\rho\|\boldsymbol{u}^{k+1}-\boldsymbol{v}^{k+1}\|_2^2+\rho\|\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\|_2^2 \tag{49}$$

where $w^k=(1/\rho)\|\boldsymbol{y}^k-\overline{\boldsymbol{y}}\|_2^2+\rho\|\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\|_2^2$. This shows that w^k monotonously decreases in each iteration until the residuals vanish. Therefore $\{(\boldsymbol{x}^k,\boldsymbol{u}^k,\boldsymbol{v}^k,\boldsymbol{y}^k)\}\to(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$. ■

APPENDIX F
PROOF OF (49)

Adding (43), (44) and (45) and multiplying by 2, we obtain

$$2\left(\boldsymbol{y}^{k+1}-\overline{\boldsymbol{y}}\right)^T\boldsymbol{r}^{k+1}+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\right)^T\left(\boldsymbol{u}^{k+1}-\overline{\boldsymbol{u}}\right)\le 0$$

where $\boldsymbol{r}^{k+1}=\boldsymbol{u}^{k+1}-\boldsymbol{v}^{k+1}$. By noticing $\boldsymbol{y}^{k+1}=\boldsymbol{y}^k+\rho\boldsymbol{r}^{k+1}$ and $\overline{\boldsymbol{u}}-\overline{\boldsymbol{v}}=\boldsymbol{0}$, we rewrite the left hand side (LHS) of the above inequality as follows:

$$\begin{aligned}\text{LHS}=&2\left(\boldsymbol{y}^k+\rho\boldsymbol{r}^{k+1}-\overline{\boldsymbol{y}}\right)^T\boldsymbol{r}^{k+1}+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\right)^T\left(\boldsymbol{r}^{k+1}+\boldsymbol{v}^{k+1}-\overline{\boldsymbol{v}}\right)\\ =&\left\{2\left(\boldsymbol{y}^k-\overline{\boldsymbol{y}}\right)^T\boldsymbol{r}^{k+1}+\rho\|\boldsymbol{r}^{k+1}\|_2^2\right\}+\left\{\rho\|\boldsymbol{r}^{k+1}\|_2^2+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\right)^T\boldsymbol{r}^{k+1}+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\right)^T\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\right)\right.\\ &\left.+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k\right)^T\left(\boldsymbol{v}^{k+1}-\overline{\boldsymbol{v}}\right)\right\}\\ =&\frac{1}{\rho}\left(\|\boldsymbol{y}^{k+1}-\overline{\boldsymbol{y}}\|_2^2-\|\boldsymbol{y}^k-\overline{\boldsymbol{y}}\|_2^2\right)+\rho\left(\|\boldsymbol{v}^{k+1}-\overline{\boldsymbol{v}}\|_2^2-\|\boldsymbol{v}^k-\overline{\boldsymbol{v}}\|_2^2\right)+\rho\|\boldsymbol{r}^{k+1}+(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k)\|_2^2\end{aligned}$$

Noticing the positivity of $(\boldsymbol{r}^{k+1})^T(\boldsymbol{v}^{k+1}-\boldsymbol{v}^k)$ according to (48), LHS ≤ 0 directly leads to (49).

REFERENCES

[1] K. Aoki, M. Fan, and A. Nishikori, "Optimal var planning by approximation method for recursive mixed-integer linear programming," *IEEE Trans. Power Syst.*, vol. 3, no. 4, pp. 1741–1747, 1988.

[2] G. Y. Yang, G. Hovland, R. Majumder, and Z. Y. Dong, "Tcsc allocation based on line flow based equations via mixed-integer programming," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2262–2269, 2007.

[3] Y.-T. Hsiao, C.-C. Liu, H.-D. Chiang, and Y.-L. Chen, "A new approach for optimal var sources planning in large scale electric power systems," *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 988–996, 1993.

[4] D. Chattopadhyay, K. Bhattacharya, and J. Parikh, "Optimal reactive power planning and its spot-pricing: an integrated approach," *IEEE Trans. Power Syst.*, vol. 10, no. 4, pp. 2014–2020, 1995.

[5] F. Dong, B. H. Chowdhury, M. L. Crow, and L. Acar, "Improving voltage stability by reactive power reserve management," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 338–345, 2005.

[6] R. Mínguez, F. Milano, R. Zarate-Miano, and A. J. Conejo, "Optimal network placement of svc devices," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1851–1860, 2007.

[7] E. Leonidaki, D. Georgiadis, and N. Hatziargyriou, "Decision trees for determination of optimal location and rate of series compensation to increase power system loading margin," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1303–1310, 2006.

[8] P. Paterni, S. Vitet, M. Bena, and A. Yokoyama, "Optimal location of phase shifters in the french network by genetic algorithm," *IEEE Trans. Power Syst.*, vol. 14, no. 1, pp. 37–42, 1999.

[9] E. Ghahremani and I. Kamwa, "Optimal placement of multiple-type facts devices to maximize power system loadability using a generic graphical user interface," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 764–778, 2013.

[10] T. Gomez, I. Perez-Arriaga, J. Lumbreras, and V. Parra, "A security-constrained decomposition approach to optimal reactive power planning," *IEEE Trans. Power Syst.*, vol. 6, no. 3, pp. 1069–1076, 1991.

[11] S. Granville and A. Lima, "Application of decomposition techniques to var planning: methodological and computational aspects," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1780–1787, 1994.

[12] F. G. Lima, F. D. Galiana, I. Kockar, and J. Munoz, "Phase shifter placement in large-scale systems via mixed integer linear programming," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1029–1034, 2003.

[13] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Foundations and Trends® in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.

[14] R. A. Jabr, N. Martins, B. C. Pal, and S. Karaki, "Contingency constrained var planning using penalty successive conic programming," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 545–553, 2012.

[15] A. Beck and Y. C. Eldar, "Sparsity constrained nonlinear optimization: Optimality conditions and algorithms," *SIAM Journal on Optimization*, vol. 23, no. 3, pp. 1480–1509, 2013.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] G. L. Torres and V. H. Quintana, "On a nonlinear multiple-centrality-corrections interior-point method for optimal power flow," *IEEE Trans. Power Syst.*, vol. 16, no. 2, pp. 222–228, 2001.

[18] Y.-C. Wu, A. S. Debs, and R. E. Marsten, "A direct nonlinear predictor-corrector primal-dual interior point algorithm for optimal power flows," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 876–883, 1994.

[19] A. Y. Yang, S. S. Sastry, A. Ganesh, and Y. Ma, "Fast l_1-minimization algorithms and an application in robust face recognition: A review," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 1849–1852.

[20] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$l_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, 2012.

[21] W. Cao, J. Sun, and Z. Xu, "Fast image deconvolution using closed-form thresholding formulas of regularization," *Journal of Visual Communication and Image Representation*, vol. 24, no. 1, pp. 31–41, 2013.

[22] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[23] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[24] Z.-B. XU, H.-L. GUO, Y. WANG, and H. ZHANG, "Representative of $l_{1/2}$ regularization among $l_q(0 < q \leq 1)$ regularizations: an experimental study based on phase diagram," *Acta Automatica Sinica*, vol. 38, no. 7, pp. 1225–1228, 2012.

[25] J. Yang and Y. Zhang, "Alternating direction algorithms for l_1-problems in compressive sensing," *SIAM journal on scientific computing*, vol. 33, no. 1, pp. 250–278, 2011.

[26] Y. Xu, W. Yin, Z. Wen, and Y. Zhang, "An alternating direction algorithm for matrix completion with nonnegative factors," *Frontiers of Mathematics in China*, vol. 7, no. 2, pp. 365–384, 2012.

[27] B. Jiang, S. Ma, and S. Zhang, "Alternating direction method of multipliers for real and complex polynomial optimization models," *Optimization*, vol. 63, no. 6, pp. 883–898, 2014.

[28] T. Erseghe, "Distributed optimal power flow using admm," *IEEE Trans. Power Syst.*, 2014.

[29] D. Phan and J. Kalagnanam, "Some efficient optimization methods for solving the security-constrained optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 863–872, March 2014.

[30] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, 2011.

[31] X.-P. Zhang, C. Rehtanz, and B. Pal, *Flexible AC transmission systems: modelling and control*. Springer, 2006.

[32] X.-P. Zhang, "Advanced modeling of the multicontrol functional static synchronous series compensator (sssc) in newton power flow," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1410–1416, 2003.

[33] S. An, J. Condren, and T. W. Gedra, "An ideal transformer upfc model, opf first-order sensitivities, and application to screening for optimal upfc locations," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 68–75, 2007.

[34] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.

[35] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 92–107, 2012.

[36] W. Bukhsh, A. Grothey, K. McKinnon, and P. Trodden, "Local solutions of the optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4780–4788, Nov 2013.

[37] F. Bach, R. Jenatton, J. Mairal, G. Obozinski *et al.*, "Structured sparsity through convex optimization," *Statistical Science*, vol. 27, no. 4, pp. 450–468, 2012.

Chao Duan (S'14) was born in Chongqing, China, in 1989. He received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China in 2012. He is currently a dual Ph.D. student at Xi'an Jiaotong University, Xi'an, China and the University of Liverpool, Liverpool, U.K. His research interests are in power system optimization and control.

Wanliang Fang was born in Henan, China, in 1958. He received the B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China in 1982 and 1988, respectively, and the Ph.D. degree from Hong Kong Polytechnic University, HongKong, in 1999, all in electrical engineering. He is currently a Professor of electrical engineering at Xi'an Jiaotong University. His research interests include power system stability analysis and control.

L. Jiang (M'00) received the B.Sc. and M.Sc. degrees from Huazhong University of Science and Technology (HUST), China, in 1992 and 1996; and the Ph.D. degree from the University of Liverpool, UK, in 2001, all in Electrical Engineering. He worked as Postdoctoral Research Assistant in the University of Liverpool from 2001 to 2003, and Postdoctoral Research Associate in the Department of Automatic Control and Systems Engineering, the University of Sheffield from 2003 to 2005. He was a Senior Lecturer at the University of Glamorgam from 2005 to 2007 and moved to the University of Liverpool at 2007. Currently, he is a Senior Lecturer in The University of Liverpool. His current research interests include control and analysis of power system, smart grid and renewable energy.

Shuanbao Niu was born in Shaanxi, China, in 1978. He received the B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China in 2000 and 2003, respectively, both in electrical engineering. He is currently a senior engineer at Northwest China Grid Company, Xi'an China. He is responsible for the analysis and control of Northwest China power system.