Appendix

A. Nomenclature

$\mathcal{N}_g, \mathcal{N}_s, \mathcal{N}_v$	Sets of generation buses, synchronous generators buses and inverters buses.
$\mathcal{N}_l, \mathcal{N}_o$	Sets of load buses and buses with neither generator nor load.
$\mathcal{N}, n_a$	The set of all buses and their number.
$\mathcal{N}_{v_o}$	The set of n_{v_o} inverters with fixed damping and inertia.
$\mathcal{N}_{v_d}$	The set of n_{v_d} inverters with adjustable damping but fixed inertia.
$\mathcal{N}_{v_m}$	The set of n_{v_m} inverters with adjustable inertia but fixed damping.
$\mathcal{N}_{v_{dm}}$	The set of $n_{v_{dm}}$ inverters with adjustable damping and inertia.
$\mathcal{B}$	The set of terminal bus pairs of all branches and $\lvert\mathcal{B}\rvert = n_b$.
$\theta \in \mathbb{R}^{n_a}$	The vector of phase angle.
$\omega \in \mathbb{R}^{n_g}$	The vector of angular frequency.
p_i	Mechanical power input for $i \in \mathcal{N}_s$, power setpoint for $i \in \mathcal{N}_v$, negative of load power independent of frequency for $i \in \mathcal{N}_l$ and 0 for $i \in \mathcal{N}_o$.
V_i	The voltage magnitude of bus i which are assumed to be constant in the study of angle stability and frequency stability.
$\underline{b}_i$	The equivalent short-circuit susceptance when ignoring the short-circuit resistance.
$\tilde{p}$	$\mathrm{col}(p_i - V_i^2 \underline{b}_i) \in \mathbb{R}^{n_a}$ with $i \in \mathcal{N}$.
M	$\mathrm{diag}(m_i) \in \mathbb{R}^{n_g \times n_g}$ with m_i being the inertia coefficient of generator i.
D	$\mathrm{diag}(d_i) \in \mathbb{R}^{n_g \times n_g}$ with d_i being the damping coefficient of generator i.
D_l	$\mathrm{diag}(d_{li}) \in \mathbb{R}^{n_l \times n_l}$ with d_{li} being the frequency coefficient of load i.
B	$\mathrm{diag}(V_{j_1} V_{j_2} b_{j_1,j_2}) \in \mathbb{R}^{n_b \times n_b}$ with $(j_1, j_2) \in \mathcal{B}$, V_i being the voltage magnitude of bus i and b_{j_1,j_2} being the susceptance of branch (j_1, j_2).
E_g, E_l, E_o, E_n	Incidence matrices showing the relationship between $\mathcal{N}_g$ and $\mathcal{N}$, $\mathcal{N}_l$ and $\mathcal{N}$, $\mathcal{N}_o$ and $\mathcal{N}$, and $\mathcal{N}$ and $\mathcal{B}$, respectively.
$\mathcal{D}$	The disturbance set.
$\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4$	Sets of power-step disturbances, power-ramp disturbances, power fluctuation disturbances and three-phase short circuit disturbances, respectively.
Superscript k	Indication of that parameters or variables correspond with disturbance k.
ρ_k	The weight coefficient of disturbance k.
W_1 *to* W_5	Diagonal weight matrices.
W_1^k	$\rho_k E_n^T W_1 E_n$.
W_2^k, W_3^k, W_4^k	$\rho_k W_2$, $\rho_k W_3$, $\rho_k W_4$.
W_5^k	$\rho_k E_l^T D_l W_5 D_l E_l$.
$\omega_{t_0} \in \mathbb{R}^{n_g}$	Initial values of ω^k.
$\theta_{t_0} \in \mathbb{R}^{n_a}$	Initial values of θ^k.
E_{gl}	$\mathrm{col}(E_g, E_l) \in \mathbb{R}^{(n_g+n_l) \times n_a}$.
$\underline{\omega}^k, \overline{\omega}^k \in \mathbb{R}^{n_g}$	Lower and upper frequency bound vectors,
$\overline{\delta} \in \mathbb{R}^{n_b}$	The upper bound vector of angle difference.
p_k	$\mathrm{col}(p_i^k) \in \mathbb{R}^{n_a}$ with p_i^k being p_i for disturbance k.
$\underline{p}_g, \overline{p}_g \in \mathbb{R}^{n_g}$	Lower and upper active power limit vectors of generators.
$\underline{m}, \overline{m} \in \mathbb{R}^{n_g}$	Lower and upper bounds of inertia coefficients.
$\underline{d}, \overline{d} \in \mathbb{R}^{n_g}$	Lower and upper bounds of damping coefficients.
$\hat{\omega}^k(t) \in \mathbb{R}^{n_g}$	The vector of piecewise polynomial with (n_c+1) degree, approximating $\omega^k(t)$.
$\hat{\theta}^k(t) \in \mathbb{R}^{n_a}$	The vector of piecewise polynomial with (n_c+1) degree, approximating $\theta^k(t)$.
Ω_i^k, Θ_i^k	Collocation coefficient matrices for profile of ω^k and θ^k at time element i.
Ω_i^k, Θ_i^k	$\mathrm{row}(\omega_{ij}^k) \in \mathbb{R}^{n_g \times (n_c+1)}$, $\mathrm{row}(\theta_{ij}^k) \in \mathbb{R}^{n_a \times (n_c+1)}$, with $j \in \{0, ..., n_c\}$.
$\boldsymbol{\omega}_i^k, \boldsymbol{\theta}_i^k$	$\mathrm{col}(\omega_{ij}^k) \in \mathbb{R}^{n_g(n_c+1)}$, $\mathrm{col}(\Theta_{ij}^k) \in \mathbb{R}^{n_a(n_c+1)}$ with $j \in \{0, ..., n_c\}$, being vector forms of collocation coefficient matrices.
$\boldsymbol{\theta}_i^{kT}, \boldsymbol{\omega}_i^{kT}$	$(\boldsymbol{\theta}_i^k)^T$, $(\boldsymbol{\omega}_i^k)^T$.
τ_j	Location of collocation points within each time element.
$\ell_\omega(\tau)$	$\ell(\tau)^T \otimes I_{n_g}$
$\ell_\theta(\tau)$	$\ell(\tau)^T \otimes I_{n_a}$.
J_k^i	The component of objective function J corresponding to disturbance k and time element i.
$\hat{J}$	The approximation of J in (P1).
δ_{jr}	The Kronecker delta.
B_{ir}^k	$B^k(t_{i-1} + \tau_r h_i^k)$.
$\tilde{p}_{ir}^k$	$\tilde{p}^k(t_{i-1} + \tau_r h_i^k)$.
$\theta_{0r}^k, \omega_{0r}^k$	Constants equal to θ_{t_0} and ω_{t_0}, respectively.
θ_{cb}	$[-\theta_c, -\theta_b, \theta_b, \theta_c]^T$.
E	$[[1,0,0],[1,1,0],[0,1,1],[0,0,1]]$.
$\check{M}, \check{D}$	$\mathrm{diag}(M, ..., M)$, $\mathrm{diag}(D, ..., D)$.
$\boldsymbol{x}$	$\mathrm{col}(M\mathbb{1}, D\mathbb{1}, \cdots, \boldsymbol{l}_{\mathrm{d}(i-1)}^k, \boldsymbol{\theta}_i^k, \boldsymbol{\omega}_i^k, \boldsymbol{l}_{\mathrm{m}i}^k, \boldsymbol{l}_{\mathrm{d}i}^k, \boldsymbol{\theta}_{i+1}^k, \cdots)$, $\forall k \in \mathcal{D}, i \in \mathbb{T}^k$.
$[\boldsymbol{x}]_i^k$	The sub-vector of $\boldsymbol{x}$ related to disturbance k and time element i, i.e., $[\boldsymbol{x}]_i^k = \mathrm{col}(M\mathbb{1}, D\mathbb{1}, \boldsymbol{\theta}_i^k, \boldsymbol{\omega}_i^k, \boldsymbol{l}_{\mathrm{m}i}^k, \boldsymbol{l}_{\mathrm{d}i}^k)$.
$[\boldsymbol{X}]_i^k$	The principal submatrix of $\boldsymbol{X}$ related to disturbance k and time element i, given as $[\boldsymbol{x}]_i^k [\boldsymbol{x}]_i^{kT}$.
$\boldsymbol{\theta}$	$\mathrm{col}(\boldsymbol{\theta}_i^k) \in \mathbb{R}^{(n_c+1)n_a \sum_{k \in \mathcal{D}} n_t^k}$ with $k \in \mathcal{D}$ and $i \in \mathbb{T}^k$.
$\boldsymbol{\omega}$	$\mathrm{col}(\boldsymbol{\omega}_i^k) \in \mathbb{R}^{(n_c+1)n_g \sum_{k \in \mathcal{D}} n_t^k}$ with $k \in \mathcal{D}$ and $i \in \mathbb{T}^k$.
$\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\varsigma}$	$\mathrm{col}(\boldsymbol{\alpha}_{(r,\imath)i}^k)$, $\mathrm{col}(\boldsymbol{\beta}_{(r,\imath)i}^k)$, $\mathrm{col}(\boldsymbol{\varsigma}_{(r,\imath)i}^k)$, with $k \in \mathcal{D}$, $i \in \mathbb{T}^k$, $r \in \{0, ..., n_c\}$, $\imath \in \mathcal{B}$, $\boldsymbol{\alpha}_{(r,\imath)i}^k \in \mathbb{R}^4$, $\boldsymbol{\beta}_{(r,\imath)i}^k \in \mathbb{R}^3$ and $\boldsymbol{\varsigma}_{(r,\imath)i}^k \in \mathbb{R}^3$.
$\tilde{\boldsymbol{Q}}_i^k$	The sub-matrix of $\boldsymbol{Q}_i^k$ by removing the last 2 block rows.
$\tilde{\boldsymbol{P}}_1$	$\frac{1}{2}[[O, I], [I, O]]$.
$\tilde{\boldsymbol{A}}_i^k, \tilde{\boldsymbol{b}}_i^k$	Sub-matrices of $\boldsymbol{A}_i^k$ and $\boldsymbol{b}_i^k$, by removing the 3th and 4th block rows, respectively.
$\boldsymbol{A}_{8i}^k$	$\mathrm{diag}(\mathrm{col}(1, 0, -1))$.

$\boldsymbol{P}^k_{(\tilde{r},\imath)i}$ With the same block structure as $\boldsymbol{P}^k_{(\hbar,r,j)i}$, which in block 3-tuple form, is given by $(\boldsymbol{\theta}^k_i, \boldsymbol{\theta}^k_i, \frac{1}{2}\vartheta\boldsymbol{A}^{kT}_{1i}O^1_{(r,j)}\boldsymbol{A}^k_{1i})$.

$O^1_{(\tilde{r},\imath)}$ $\in \mathbb{R}^{n_b(n_c+1)\times n_b(n_c+1)}$, being a matrix with diagonal elements corresponding to the rth time element and branch $\imath$ being 1 and others being 0.

$\boldsymbol{A}^k_{9i}$ $\mathrm{diag}(\mathrm{col}(\pi\vartheta, \frac{\sin\theta_b}{\theta_b}, \pi\vartheta))\boldsymbol{A}_{1i}$.

$\boldsymbol{A}^k_{10i}$ $\mathrm{diag}(\theta^T_{cb})$.

$\tilde{\boldsymbol{P}}_2$ $[[I,O],[O,O]]$.

$\mathbb{S}^{|\mathcal{V}|}_{\mathcal{E}}$ The set of symmetric $|\mathcal{V}|\times|\mathcal{V}|$ matrices with only entries in $\mathcal{E}$ specified.

$\mathcal{S}_{\mathcal{C}}(\boldsymbol{Z})$ The principal submatrix of $\boldsymbol{Z}$ defined by the index set $\mathcal{C}$ or a new defined matrix for the principal submatrix.

$\mathcal{I}(\cdot,\cdot)$ The set of row (or column) index of $\boldsymbol{Z}$ corresponding to that in parentheses. For example, $\mathcal{I}(\boldsymbol{\theta}^k_i,\cdot)$ denotes the set of row index of $\boldsymbol{Z}$ corresponding to $\boldsymbol{\theta}^k_i$, $\mathcal{I}(\boldsymbol{\theta}^k_i, j)$ denotes the set of row index of $\boldsymbol{Z}$ corresponding to $\boldsymbol{\theta}^k_i$ and bus (or generator) j or collocation point j, $\mathcal{I}(-1)$ represents the index of the last row of $\boldsymbol{Z}$, and $\mathcal{I}(k,i)$ represents the set of row index of $\boldsymbol{Z}$ only corresponding to disturbance k and time element i,i.e., $\mathcal{I}(k,i)=\mathcal{I}(\boldsymbol{\theta}^k_i,\cdot)\cup\mathcal{I}(\boldsymbol{\omega}^k_i,\cdot)\cup\mathcal{I}(\boldsymbol{l}^k_{\mathrm{m}i},\cdot)\cup\mathcal{I}(\boldsymbol{l}^k_{\mathrm{d}i},\cdot)$.

$\tilde{\mathcal{G}}(\tilde{\mathcal{V}},\tilde{\mathcal{E}})$ The associated graph of $\tilde{\boldsymbol{Z}}$.

$\tilde{\mathcal{I}}^k_i(\boldsymbol{\beta},r,\imath,j)$ The index of $\tilde{\boldsymbol{Z}}$ corresponding to the jth entries in $\boldsymbol{\beta}^k_{(r,\imath)i}$, with $(r,\imath,j)\in\mathcal{P}=\{0,...,n_c\}\times\mathcal{B}\times\{1,2,3\}$.

$\tilde{\mathcal{I}}^k_i(\boldsymbol{\varsigma},r,\imath,j)$ The index of $\tilde{\boldsymbol{Z}}$ corresponding to the jth entries in $\boldsymbol{\varsigma}^k_{(r,\imath)i}$, with $(r,\imath,j)\in\mathcal{P}=\{0,...,n_c\}\times\mathcal{B}\times\{1,2,3\}$.

$\mathbf{ch}(\mathcal{C}_{ej})$ The set of children of clique $\mathcal{C}_{ej}$ in clique tree $\mathcal{T}$.

$\mathcal{K}^j$ $\{\mathcal{C}_{ej}|\mathcal{C}_{ej}\in\bigcup_{k\in\mathcal{D},i\in\mathbb{T}^k}\mathcal{K}^{ki}_{e1}\}$ with $j\in\mathcal{N}_g$.

I_j, I'_j Sets of indices corresponding to the intersection between cliques $\mathcal{C}_{ej}$ and $\mathcal{C}'_{ej}$, for $\hat{\boldsymbol{Z}}_{\mathcal{C}_{ej}}$ and $\hat{\boldsymbol{Z}}_{\mathcal{C}'_{ej}}$, respectively.

I_ω The set of indices consisting of -1 and the index corresponding to ω^k_{i0} of generator j.

I'_ω The set of indices consisting of -1 and the index corresponding to ω^k_{i-1,n_c} of generator j.

I_θ The set of indices consisting of -1 and the index corresponding to θ^k_{i0} of buses $\mathcal{C}_{nj}$.

I'_θ The set of indices consisting of -1 and the index corresponding to θ^k_{i-1,n_c} of buses $\mathcal{C}_{nj}$.

A_{adj} The adjacent matrix of $\mathcal{G}_n$.

β_{cf} A proper large number to guarantee positive definiteness of $A_{adj}+\beta_{cf}I$.

$\boldsymbol{P}_{\mathcal{C}_{ej}}, \boldsymbol{P}^\gamma_{\mathcal{C}_{ej}}$ Proper matrices to make $\mathrm{Tr}(\boldsymbol{P}^k_{0i}[\boldsymbol{X}]^k_i) = \sum_{\mathcal{C}_{ej}\in\mathcal{K}^k_i\cup\tilde{\mathcal{K}}^k_i}\mathrm{Tr}(\boldsymbol{P}_{\mathcal{C}_{ej}}\hat{\boldsymbol{Z}}_{\mathcal{C}_{ej}})$, constraints (19b) equivalent to (13b), (13c), (16a) to the first equation of (16d) and (16e) to (16h), respectively.

$\max^\circ, \min^\circ$ Return maximum and minimum values of each continuous part of an integer set, respectively, e.g., $\max^\circ\{1,2,4,5,9\}=\{2,5,9\}$ and $\min^\circ\{1,2,4,5,9\}=\{1,5,9\}$.

$\Im(\Xi_s)$ Return an arbitrary subset of Ξ_s with one element.

$\tilde{\mathcal{M}}$ $\bigcup_{s\in\mathbb{P}}\tilde{\mathcal{M}}_s$.

$\mathcal{O}_s$ A proper-sized matrix vector consisting of zero matrices.

$\mathrm{diag}(\zeta)_i$ n_{d} arbitrary disjoint sub-vectors of $\mathrm{diag}(\zeta)$ satisfying $[\mathrm{diag}(\zeta)^T_1,\mathrm{diag}(\zeta)^T_2,...,\mathrm{diag}(\zeta)^T_{n_{\mathrm{d}}}]=\mathrm{diag}(\zeta)^T$.

$\mathrm{upper}(\zeta)_i$ Analogous to $\mathrm{diag}(\zeta)_i$.

$\mathcal{O}'_s$ A matrix vector consisting of zero matrices, with the same size as $\mathcal{Y}_s$.

$\tilde{\rho}>0$ The penalty parameter.

$\sigma^1_{\mathcal{C}_{ej}}, u^1_{\mathcal{C}_{ej}}, v^1_{\mathcal{C}_{ej}}$ The 1-th singular value, left singular vector and right singular vector of matrix $\hat{\boldsymbol{Z}}^{(\kappa+1)}_{\mathcal{C}_{ej}} + \frac{1}{\tilde{\rho}}\hat{\boldsymbol{\Lambda}}^{(\kappa)}_{\mathcal{C}_{ej}}$, respectively.

$\epsilon^{\mathrm{abs}}, \epsilon^{\mathrm{rel}}$ Absolute tolerance and relative tolerance.

$\boldsymbol{Z}^{\mathbf{md}}_{j,(1,2)}$ The entry in the 1-th row and 2-th column of $\boldsymbol{Z}^{\mathrm{md}}_j$.

$\varphi_s, \tilde{\varphi}_s$ Slack variables.

$\overline{\mathcal{M}}_s$ $\{(k,i_k)|(k,i_k)\in\Xi_s\wedge i_k\in\max^\circ\{i|(k,i)\in\Xi_s\}\backslash\{n^k_t\}\}, \forall s\in\mathbb{P}$.

$\underline{\mathcal{M}}_s$ $\{(k,i_k)|(k,i_k)\in\Xi_s\wedge i_k\in\min^\circ\{i|(k,i)\in\Xi_s\}\backslash\{1\}\}, \forall s\in\mathbb{P}$.

$\underline{\mathcal{M}}$ $\bigcup_{s\in\mathbb{P}}\underline{\mathcal{M}}_s$.

$\hat{\mathcal{Z}}_s$ $[\hat{\boldsymbol{Z}}_{\mathcal{C}_{ej}}]^T$ with $\mathcal{C}_{ej}\in\mathcal{K}^k_i$ and $(k,i)\in\Xi_s$.

$\mathbb{Z}_s$ The feasible region defined by inner constraints with $(k,i)\in\Xi_s$ and coupling constraints $(k,i)\in\Xi_s$ satisfying $(k,i-1)\in\Xi_s$ or $i=1$.

$\mathcal{Z}_{\mathbf{a}}$ The matrix vector consisting of all auxiliary matrix variables.

$\mathbb{R}_{\mathbf{a}}$ The feasible region of $\mathcal{Z}_{\mathrm{a}}$.

B. Derivation of NLP formulation of DID

Lagrange polynomial ℓ_j satisfies $\ell_j(\tau_r)=\delta_{jr}$ for $\forall j,r\in\{0,...,n_c\}$, with δ_{jr} being the Kronecker delta. $\hat{\omega}^k$ and $\hat{\theta}^k$ have the property that $\hat{\omega}^k(t_{i-1}+\tau_j h^k_i)=\omega^k_{ij}$ and $\hat{\theta}^k(t_{i-1}+\tau_j h^k_i)=\theta^k_{ij}$, respectively. With this property, substituting the polynomial into DAE constraints (2b) and enforcing the resulting algebraic equations at the interpolation points τ_r lead to the following collocation equations for DAE constraints:

$$\begin{cases} E_g\dot{\ell}_\theta(\tau_r)\boldsymbol{\theta}^k_i = h^k_i\omega^k_{ir} \quad \forall r\in\{1,...,n_c\} \\ \dot{\ell}_\omega(\tau_r)\boldsymbol{\omega}^k_i = -h^k_i M^{-1}D\omega^k_{ir} + h^k_i M^{-1}E_g\tilde{p}^k_{ir} \\ \qquad - h^k_i M^{-1}E_gE_nB^k_{ir}\sin(E^T_n\theta^k_{ir}) \quad \forall r\in\{1,...,n_c\} \\ E_l\dot{\ell}_\theta(\tau_r)\boldsymbol{\theta}^k_i = h^k_i D^{-1}_l E_l\tilde{p}^k_{ir} \\ \qquad - h^k_i D^{-1}_l E_lE_nB^k_{ir}\sin(E^T_n\theta^k_{ir}) \quad \forall r\in\{1,...,n_c\} \\ 0 = E_o\tilde{p}^k_{ir} - E_oE_nB^k_{ir}\sin(E^T_n\theta^k_{ir}) \quad \forall r\in\{0,...,n_c\} \end{cases} \tag{38}$$

where $B^k_{ir}=B^k(t_{i-1}+\tau_r h^k_i)$ and $\tilde{p}^k_{ir}=\tilde{p}^k(t_{i-1}+\tau_r h^k_i)$.

Additionally, ω^k_{i0} and $E_{gl}\theta^k_{i0}$ are determined by initial conditions (2c), or enforced by the continuity of the differential variable profiles across element boundaries as follows:

$$\begin{bmatrix}\omega_{i0}^k\\ E_{gl}\theta_{i0}^k\end{bmatrix}=\begin{cases}\begin{bmatrix}\omega_{t_0}\\ E_{gl}\theta_{t_0}\end{bmatrix} & i=1\\ \begin{bmatrix}\hat{\omega}^k(t_{i-2}+\tau_{n_c}h_{i-1}^k)=\omega_{(i-1)n_c}^k\\ E_{gl}\hat{\theta}^k(t_{i-2}+\tau_{n_c}h_{i-1}^k)=E_{gl}\theta_{(i-1)n_c}^k\end{bmatrix} & i\in\{2,\ldots,n_t^k\}\end{cases} \tag{39}$$

Path constraints (2d) to (2e) are enforced at all collocation points, then we have:

$$\underline{\omega}_{ir}^k \le \omega_{ir}^k \le \overline{\omega}_{ir}^k \qquad \forall r\in\{1,\ldots,n_c\} \tag{40}$$

$$-\overline{\delta} \le \theta_{ir}^k \le \overline{\delta} \qquad \forall r\in\{1,\ldots,n_c\} \tag{41}$$

$$\underline{p}_g \le E_g p_{ir}^k - \frac{1}{h_i^k} M\dot{\ell}_\theta(\tau_r)\boldsymbol{\omega}_i^k - D\omega_{ir}^k \le \overline{p}_g \quad \forall r\in\{1,\ldots,n_c\} \tag{42}$$

where $\underline{\omega}_{ir}^k=\underline{\omega}^k(t_{i-1}+\tau_r h_i^k)$, $\overline{\omega}_{ir}^k=\overline{\omega}^k(t_{i-1}+\tau_r h_i^k)$ and $p_{ir}^k=p^k(t_{i-1}+\tau_r h_i^k)$.

Finally, combining (2f), (9), and 38 to (42), we can obtain the NLP formulation of DID.

C. Matrices in the NLP formulation of DID: See Table II.

D. Matrices in the QCQP formulation of DID: See Table III.

TABLE III
MATRICES IN THE QCQP FORMULATION OF DID

$\boldsymbol{P}_{0i}^k$	$\boldsymbol{b}_i^k$
$\begin{bmatrix}\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{W}_{1i}^k&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{W}_{2i}^k&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{W}_{3i}^k&\frac{1}{2}\boldsymbol{W}_{4i}^k\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\frac{1}{2}\boldsymbol{W}_{4i}^k&\boldsymbol{W}_{5i}^k\end{bmatrix}$	$\begin{bmatrix}-(\boldsymbol{b}_{2i}^k+\boldsymbol{C}_{2i}^k\boldsymbol{\Lambda}_{1i}^k)\\ \boldsymbol{b}_{2i}^k+\boldsymbol{C}_{2i}^k\boldsymbol{\Lambda}_{1i}^k\\ \overline{\boldsymbol{c}}_{1i}^k+\boldsymbol{b}_{1i}^k+\boldsymbol{C}_{1i}^k\boldsymbol{\Lambda}_{1i}^k\\ -(\underline{\boldsymbol{c}}_{1i}^k+\boldsymbol{b}_{1i}^k+\boldsymbol{C}_{1i}^k\boldsymbol{\Lambda}_{1i}^k)\\ \overline{\boldsymbol{c}}_{2i}^k\\ -\underline{\boldsymbol{c}}_{2i}^k\end{bmatrix}$
$\boldsymbol{Q}_i^k$	$\tilde{\boldsymbol{b}}_{1i}^k$
$\begin{bmatrix}\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{C}_{1i}^k\boldsymbol{\Lambda}_{0i}^k&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&-I_{n_g(n_c+1)}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}&-I_{n_g(n_c+1)}\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{b}_{1i}^k+\boldsymbol{C}_{1i}^k\boldsymbol{\Lambda}_{1i}^k\\ \boldsymbol{O}\\ \boldsymbol{O}\end{bmatrix}$

$\boldsymbol{A}_i^k$	$\boldsymbol{L}_{13}^T$	$\boldsymbol{L}_{14}^T$
$\begin{bmatrix}\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{C}_{2i}^k\boldsymbol{\Lambda}_{0i}^k+\boldsymbol{A}_{2i}^k&\boldsymbol{A}_{3i}^k&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&-(\boldsymbol{C}_{2i}^k\boldsymbol{\Lambda}_{0i}^k+\boldsymbol{A}_{2i}^k)&-\boldsymbol{A}_{3i}^k&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&-\boldsymbol{C}_{1i}^k\boldsymbol{\Lambda}_{0i}^k&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}&\boldsymbol{C}_{1i}^k\boldsymbol{\Lambda}_{0i}^k&\boldsymbol{O}&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{A}_{6i}^k&\boldsymbol{A}_{7i}^k&\boldsymbol{A}_{5i}^k&\boldsymbol{A}_{4i}^k&\boldsymbol{O}&\boldsymbol{O}\\ -\boldsymbol{A}_{6i}^k&-\boldsymbol{A}_{7i}^k&-\boldsymbol{A}_{5i}^k&-\boldsymbol{A}_{4i}^k&\boldsymbol{O}&\boldsymbol{O}\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{L}_1&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{L}_3\\ \boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{L}_2&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{L}_4\\ \boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}\end{bmatrix}$
$\boldsymbol{P}_{(1,r,j)i}^k=(M\mathbb{1},\boldsymbol{\omega}_i^k,\frac{1}{2}\boldsymbol{\ell}_{\omega j}(\tau_r)),\ (D\mathbb{1},\boldsymbol{\omega}_i^k,\frac{1}{2}E_{(r,j)})$		
$\boldsymbol{P}_{(2,r,j)i}^k=(M\mathbb{1},\boldsymbol{\omega}_i^k,\frac{1}{2}\boldsymbol{O}_{(r,j)}^1)$, $\boldsymbol{P}_{(3,r,j)i}^k=(D\mathbb{1},\boldsymbol{\omega}_i^k,\frac{1}{2}\boldsymbol{O}_{(r,j)}^1)$		

Note: $\boldsymbol{\Lambda}_{0i}^k=\operatorname{diag}\left(\cos(\boldsymbol{A}_{1i}^k\boldsymbol{\theta}_0^k)\right)$ and $\boldsymbol{\Lambda}_{1i}^k=\sin(\boldsymbol{A}_{1i}^k\boldsymbol{\theta}_0^k)-\boldsymbol{\Lambda}_{0i}^k\boldsymbol{\theta}_0^k$. Since $\boldsymbol{P}_{(\hbar,r,j)i}^k$ are highly sparse and symmetric matrices, they are given in block 3-tuple form and only with lower triangular portions of matrix. $\boldsymbol{\ell}_{\omega j}(\tau_r))$ equals to $\boldsymbol{\ell}_\omega(\tau_r))$ with only the row corresponding to d_j remained and others replaced by 0. $\boldsymbol{O}_{(r,j)}^1\in\mathbb{R}^{n_g\times n_g(n_c+1)}$ is a matrix with the element corresponding to d_j and angular speed of generator j in $\boldsymbol{\omega}_{ir}^k$ being 1 and other being 0.

TABLE II
MATRICES IN THE NLP FORMULATION OF DID

$\boldsymbol{A}_{2i}^k$	$\boldsymbol{A}_{3i}^k$	$\boldsymbol{A}_{4i}^k$	$\boldsymbol{A}_{5i}^k$	$\boldsymbol{A}_{6i}^k$	$\boldsymbol{A}_{7i}^k$
$\begin{bmatrix}\operatorname{col}(E_g\dot{\ell}_\theta(\tau_{r_1}))\\ \operatorname{col}(D_lE_l\dot{\ell}_\theta(\tau_{r_1}))\\ O\end{bmatrix}$	$\begin{bmatrix}-E_2h_i^k\\ O\\ O\end{bmatrix}$	$\begin{bmatrix}E_2\\ O\\ O\\ O\end{bmatrix}$	$\begin{bmatrix}O\\ E_3\\ O\\ O\end{bmatrix}$	$\begin{bmatrix}O\\ O\\ I_{n_g}\\ O\end{bmatrix}$	$\begin{bmatrix}O\\ O\\ O\\ I_{n_g}\end{bmatrix}$

$\boldsymbol{A}_{1i}^k$	$\boldsymbol{H}_{1i}^k$	$\boldsymbol{H}_{2i}^k$	$\boldsymbol{C}_{1i}^k$
$\operatorname{diag}(E_n^T)$	$\operatorname{col}(\dot{\ell}_\omega(\tau_{r_1}))$	$h_i^kE_2$	$h_i^k\operatorname{diag}(E_gE_nB_{ir_1}^k)E_1$

$\boldsymbol{C}_{2i}^k$	$\boldsymbol{b}_{2i}^k$	$\underline{\boldsymbol{c}}_{2i}^k$	$\overline{\boldsymbol{c}}_{2i}^k$
$\begin{bmatrix}O\\ h_i^k\operatorname{diag}(E_lE_nB_{ir_1}^k)E_1\\ -\operatorname{diag}(E_oE_nB_{ir_0}^k)E_1\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{0}\\ -\operatorname{col}(h_i^kE_l\tilde{p}_{ir_1}^k)\\ \operatorname{col}(E_o\tilde{p}_{ir_0}^k)\end{bmatrix}$	$\begin{bmatrix}\operatorname{col}(\underline{\omega}_{ir_1}^k)\\ \operatorname{col}(-\overline{\delta})\\ \underline{m}\\ \underline{d}\end{bmatrix}$	$\begin{bmatrix}\operatorname{col}(\overline{\omega}_{ir_1}^k)\\ \operatorname{col}(\overline{\delta})\\ \overline{m}\\ \overline{d}\end{bmatrix}$

$\boldsymbol{b}_{1i}^k$	$\underline{\boldsymbol{c}}_{1i}^k$	$\overline{\boldsymbol{c}}_{1i}^k$
$-\operatorname{col}(h_i^kE_g\tilde{p}_{ir_1}^k)$	$h_i^k\operatorname{col}(E_gp_{ir_1}^k-\overline{p}_g)$	$h_i^k\operatorname{col}(E_gp_{ir_1}^k-\underline{p}_g)$
$\boldsymbol{L}_1$	$\boldsymbol{L}_2$	$\boldsymbol{L}_3$
$E_{gl}\operatorname{row}(I_{n_a},O_{n_a\times n_an_c})$	$E_{gl}\operatorname{row}(O_{n_a\times n_an_c},I_{n_a})$	$\operatorname{row}(I_{n_g},O_{n_g\times n_gn_c})$
$\boldsymbol{L}_4$	E_1	E_2
$\operatorname{row}(O_{n_g\times n_gn_c},I_{n_g})$	$\operatorname{row}(O_{n_bn_c\times n_b},I_{n_bn_c})$	$\operatorname{row}(O_{n_gn_c\times n_g},I_{n_gn_c})$
E_3	r_0	r_1
$\operatorname{row}(O_{n_bn_c\times n_a},\operatorname{diag}(E_n^T))$	$\{0,\ldots,n_c\}$	$\{1,\ldots,n_c\}$

$\boldsymbol{W}_{1i}^k$	$\boldsymbol{W}_{2i}^k$
$h_i^k(S_1\otimes W_2^k)+\frac{1}{h_i^k}(S_2\otimes W_3^k)$	$\frac{1}{h_i^k}(S_2\otimes W_4^k)$

$\boldsymbol{W}_{3i}^k$	$\boldsymbol{W}_{4i}^k$	$\boldsymbol{W}_{5i}^k$
$h_i^k(S_1\otimes W_1^k)+\frac{1}{h_i^k}(S_2\otimes W_5^k)$	$2(S_3\otimes W_4^k)$	$h_i^k(S_1\otimes W_4^k)$
S_1	S_2	S_3
$\int_0^1\ell(\tau)\ell(\tau)^T\mathrm{d}\tau$	$\int_0^1\dot{\ell}(\tau)\dot{\ell}(\tau)^T\mathrm{d}\tau$	$\int_0^1\dot{\ell}(\tau)\ell(\tau)^T\mathrm{d}\tau$

E. Approximation 1

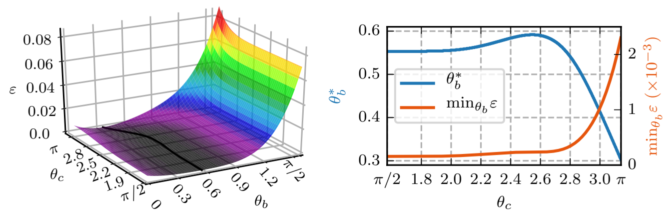
Approximation 1. $\sin\theta\mapsto\beta^T\varsigma$ *with*

$$\begin{bmatrix}\vartheta\theta^2+\pi\vartheta\theta+\frac{\pi^2}{4}\vartheta-1\\ \frac{\sin\theta_b}{\theta_b}\theta\\ -\vartheta\theta^2+\pi\vartheta\theta-\frac{\pi^2}{4}\vartheta+1\end{bmatrix}-\varsigma=0 \textit{ with } \vartheta=\frac{1-\sin\theta_b}{(\theta_b-\frac{\pi}{2})^2},$$

$$\theta-\theta_{cb}^T\alpha=0, \mathbb{1}^T\beta=1, \mathbb{1}^T\alpha=1, (\beta-\mathbb{1})^2=0, \alpha-E\beta\le 0, \alpha\ge 0$$

where $\theta\in[-\theta_c,\theta_c]$ *with* $\theta_c\in[\frac{\pi}{2},\pi]$, $\beta\in\mathbb{R}^3$, $\varsigma\in\mathbb{R}^3$, $\alpha\in\mathbb{R}^4$; $\theta_b\in[0,\frac{\pi}{2}]$ *and* θ_c *are known parameters.*

The essence of Approximation 1 is using quadratic function ς_1, linear function ς_2 and quadratic function ς_3 to approximate $\sin\theta$ for $\theta\in[-\theta_c,-\theta_b]$, $\theta\in[-\theta_b,\theta_b]$ and $\theta\in[\theta_b,\theta_c]$, respectively. Clearly, θ_b can observably impact approximation errors and should be selected carefully. Define approximation error function $\epsilon:\theta_b\to\int_{-\theta_c}^{\theta_c}(\sin\theta-\beta^T\varsigma)^2\mathrm{d}\theta$ and function $\theta_b^*:\theta_c\to\{\arg\min_{\theta_b}\epsilon, \text{s.t.}\,\theta_b\in[0,\frac{\pi}{2}]\}$. Graphs of θ_b-θ_c-ϵ, θ_c-θ_b^* and θ_c-$\min_{\theta_b}\epsilon$ are shown in Fig. 5. Numerically, we can find that $\forall\theta_c\in[\frac{\pi}{2},\pi]$, ϵ is a convex function in domain $[0,\frac{\pi}{2}]$.



(a) Function surface of θ_c-θ_b-ϵ (b) Curves of θ_c-θ_b^* and θ_c-$\min_{\theta_b}\epsilon$

Fig. 5. Numerical analysis of Approximation 2.

Thus, given $\theta_c\in[\frac{\pi}{2},\pi]$, by solving $\{\arg\min_{\theta_b}\epsilon,\text{s.t.}\theta_b\in[0,\frac{\pi}{2}]\}$, the unique optimal value of $\theta_b=\theta_b^*$ to minimize the approximation error can be obtained. Numerically, it can be found that for $\theta_c\in[\frac{\pi}{2},\frac{9}{10}\pi]$, Approximation 1 can be with very high accuracy.

F. Analysis of the aggregate sparsity pattern of $\mathbf{Z}$

Fig. 6(a) shows the aggregate sparsity pattern of matrix $\boldsymbol{Z}$ at the block level with $\boldsymbol{Z}$ broken into blocks corresponding to each pair of disturbances and time elements. We can find that the sparsity graph is chordal and its set of maximal cliques is given by:

$$\mathcal{K}_b=\{\mathcal{C}_i^k|k\in\mathcal{D},i\in\mathbb{T}^k,\mathcal{C}_i^k=\{\mathcal{I}(M\mathbb{1},\cdot),\mathcal{I}(D\mathbb{1},\cdot),\mathcal{I}(k,i),\mathcal{I}(-1)\}\} \tag{43}$$

Note that dashed edges in Fig. 6(a) are extra added to reduced the number of equality constraints for overlapping entries. Otherwise, $\forall k\in\mathcal{D},i\in\mathbb{T}^k$, equality constraints for some entries in $\mathcal{I}(k,i)$ have to be introduced. Furthermore, Fig. 6(b) shows the aggregate sparsity pattern of $\boldsymbol{Z}$ within each clique of $\mathcal{K}_b$, still at the block level. The sparsity graph is also chordal and its set of maximal cliques is given by:

$$\begin{aligned}\mathcal{K}_b^{ki}=\{&\mathcal{C}_{bj}|j=1,2,3,\mathcal{C}_{b1}=\{\mathcal{I}(M\mathbb{1},\cdot),\mathcal{I}(D\mathbb{1},\cdot),\mathcal{I}(\boldsymbol{\omega}_i^k,\cdot),\mathcal{I}(-1)\},\\&\mathcal{C}_{b2}=\{\mathcal{I}(\boldsymbol{l}_{\text{mi}}^k,\cdot),\mathcal{I}(\boldsymbol{l}_{\text{di}}^k,\cdot),\mathcal{I}(-1)\},\mathcal{C}_{b3}=\{\mathcal{I}(\boldsymbol{\theta}_i^k,\cdot),\mathcal{I}(-1)\}\}\end{aligned} \tag{44}$$

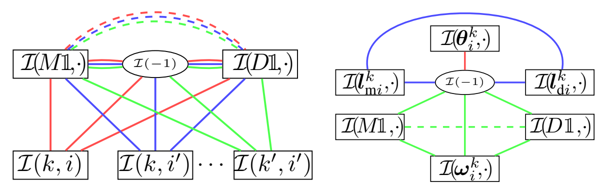
Then we investigate aggregate sparsity patterns at the element level for each clique in $\mathcal{K}_b^{ki}$ individually. For cliques $\mathcal{C}_{b1}$ and $\mathcal{C}_{b2}$, they have an analogous aggregate sparsity pattern at the element level. Taking a power grid with 3 generators and n_c=1 as an example, Fig. 7(a) shows the aggregate sparsity pattern within $\mathcal{C}_{b1}$ or $\mathcal{C}_{b2}$. Nodes with the same color are associated with indices corresponding to the same generator. It can be found that the sparsity graph is chordal, and each set of nodes corresponding to the same generator and node $\mathcal{I}(-1)$ consist of a maximal clique. Thus sets of maximal cliques for $\mathcal{C}_{b1}$ and $\mathcal{C}_{b2}$ are given as

$$\begin{aligned}&\mathcal{K}_{e1}^{ki}=\{\mathcal{C}_{ej}|\mathcal{C}_{ej}=\mathcal{I}(M\mathbb{1},j)\cup\mathcal{I}(D\mathbb{1},j)\cup\mathcal{I}(\boldsymbol{\omega}_i^k,j)\cup\mathcal{I}(-1),j\in\mathcal{N}_g\},\\&\mathcal{K}_{e2}^{ki}=\{\mathcal{C}_{ej}|\mathcal{C}_{ej}=\mathcal{I}(\boldsymbol{l}_{\text{mi}}^k,j)\cup\mathcal{I}(\boldsymbol{l}_{\text{di}}^k,j)\cup\mathcal{I}(-1),j\in\mathcal{N}_g\}\end{aligned} \tag{45}$$

respectively.

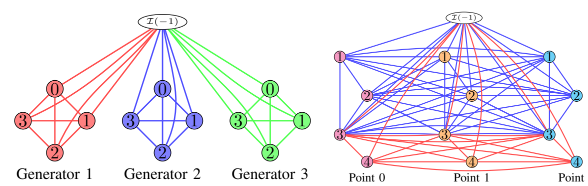
For clique $\mathcal{C}_{b3}$, Fig. 7(b) gives an example of its aggregate sparsity pattern, where the power grid consists of 4 buses and n_c=2. Nodes with the same color are associated with indices corresponding to the same collocation point. The induced subgraph for any set of nodes with the same color is the same as the underlying graph of the power grid. With numbers in colored nodes representing bus numbers of power grid, the underlying graph of power grid contains two maximal cliques, i.e., $\{1,2,3\}$ and $\{3,4\}$. Node $\mathcal{I}(-1)$ and all nodes that correspond to each maximal clique of the underlying graph of power grid, form a maximal clique of the sparsity graph of $\mathcal{C}_{b3}$. These two maximal cliques are shown in Fig. 7(b) as the two groups of nodes linked by blue edges and red edges, respectively. We can find that the aggregate sparsity pattern within $\mathcal{C}_{b3}$ is fully determined by the topology of power grids. Specifically, denoting the underlying graph of power grid by $\mathcal{G}_n(\mathcal{N},\mathcal{B})$, the sparsity graph of $\mathcal{C}_{b3}$ is chordal if and only if $\mathcal{G}_n$ is chordal. Without loss of generality, it is assumed that $\mathcal{G}_n$ is chordal with its set of maximal cliques denoted by $\mathcal{K}_n$=$\{\mathcal{C}_{nj}\}$, since we can always find a chordal extension for $\mathcal{G}_n$. Then the set of maximal cliques for $\mathcal{C}_{b3}$ is given by

$$\mathcal{K}_{e3}^{ki}=\{\mathcal{C}_{ej}|\mathcal{C}_{ej}=\mathcal{I}(\boldsymbol{\theta}_i^k,\mathcal{C}_{nj})\cup\mathcal{I}(-1),\mathcal{C}_{nj}\in\mathcal{K}_n\} \tag{46}$$



(a) (b)

Fig. 6. Aggregate sparsity pattern of matrix $\boldsymbol{Z}$ at the block level.



(a) (b)

Fig. 7. Aggregate sparsity pattern of matrix $\boldsymbol{Z}$ at the element level.

G. Matrix vector and its operations

We call $\mathcal{X}$=$[\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_n]^T$ where $\boldsymbol{X}_i$ are all matrices, a *matrix vector*. If another matrix vector $\mathcal{Y}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2,...,\boldsymbol{Y}_n]^T$ satisfies that $\forall i=1,2,...,n$, $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ are with the same size, then $\mathcal{X}$ and $\mathcal{Y}$ are with the same size. $\mathcal{X}\pm\mathcal{Y}=[\boldsymbol{X}_1\pm\boldsymbol{Y}_1,...,\boldsymbol{X}_n\pm\boldsymbol{Y}_n]^T$. $\mathcal{X}^{T^*}\circ\mathcal{Y}=[\boldsymbol{X}_1^T\boldsymbol{Y}_1,\boldsymbol{X}_2^T\boldsymbol{Y}_2,...,\boldsymbol{X}_n^T\boldsymbol{Y}_n]^T$. $\text{Tr}(\mathcal{X})$=$\sum_{i=1}^n\text{Tr}(\boldsymbol{X}_i)$. $\mathcal{X}^{T^*}$=$[\boldsymbol{X}_1^T,\boldsymbol{X}_2^T,...,\boldsymbol{X}_n^T]^T$. Frobenius norm of $\mathcal{X}$ is define as: $\|\mathcal{X}\|_F=\sqrt{\sum_{i=1}^n\text{Tr}(\boldsymbol{X}_i^T\boldsymbol{X}_i)}$. A *linear matrix vector function* is defined as: $\mathcal{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_n]^T\rightarrow\mathcal{Z}=[\boldsymbol{Z}_1,\boldsymbol{Z}_2,...,\boldsymbol{Z}_m]^T$, where $\mathcal{X}$ and $\mathcal{Z}$ are matrix vectors, and $\forall j=1,2,...,m$, $\exists i=1,2,...,n$ and matrix $\boldsymbol{P}_j$, let $\boldsymbol{Z}_j=\boldsymbol{P}_j\boldsymbol{X}_i\boldsymbol{P}_j^T$. $\text{rank}(\mathcal{X})$=$\max\{\text{rank}(\boldsymbol{X}_i)|i$=$1,2,..,n\}$. $\text{vec}(\mathcal{X})=[\text{vec}(\boldsymbol{X}_1)^T,\ \text{vec}(\boldsymbol{X}_2)^T,....,\text{vec}(\boldsymbol{X}_n)^T]^T$ with $\text{vec}(\boldsymbol{X}_i)$ being the vectorization of matrix $\boldsymbol{X}_i$. $\mathcal{X}$ is called a symmetric matrix vector or $\mathcal{X}$ is symmetric if $\forall i$=$1,2,...,n$, $\boldsymbol{X}_i$ is a symmetric matrix. $\text{diag}(\mathcal{X})$=$[\text{diag}(\boldsymbol{X}_1)^T,\text{diag}(\boldsymbol{X}_2)^T,...,\text{diag}(\boldsymbol{X}_n)^T]^T$. $\text{lower}(\mathcal{X})$=$[\text{lower}(\boldsymbol{X}_1)^T,\text{lower}(\boldsymbol{X}_2)^T,...,\text{lower}(\boldsymbol{X}_n)^T]^T$ with lower$(\boldsymbol{X}_i)^T$ being the vectorization of all the entries under the main diagonal in matrix $\boldsymbol{X}_i$ following a column-major order. upper$(\mathcal{X})$ is analogous to lower$(\mathcal{X})$ but for all the entries above the main diagonal and following a row-major order.

H. Size reduction for PSD constraints

The size of PSD constraint (25) can be reduced by considering the symmetry of $\zeta^A_s(\hat{\mathcal{Z}}_s) - (\zeta^B_s(\mathcal{Z}^{(\kappa)}_{\mathrm{a}}) - \frac{1}{\rho}\mathcal{A}^{(\kappa)}_s)$ and introducing multiple slack variables. For simplicity, we use ζ to represent $\zeta^A_s(\hat{\mathcal{Z}}_s) - (\zeta^B_s(\mathcal{Z}^{(\kappa)}_{\mathrm{a}}) - \frac{1}{\rho}\mathcal{A}^{(\kappa)}_s)$ in this remark only. In constraint (25), $\zeta^A_s(\hat{\mathcal{Z}}_s)$ and $\zeta^B_s(\mathcal{Z}^{(\kappa)}_{\mathrm{a}})$ are symmetric for all $\kappa \geq 0$. Thus ζ is symmetric as long as $\mathcal{A}^{(\kappa)}_s$ is symmetric, which can be guaranteed by setting $\mathcal{A}^{(0)}_s$ to a symmetric matrix vector. Furthermore, upper(ζ) = lower(ζ) with ζ being symmetric. Therefor, if $\mathcal{A}^{(0)}_s$ is symmetric, constraint (25) can be replace by the following equivalent form:

$$\begin{cases} \varphi_s = 2\sum_{i=1}^{n_{\mathrm{u}}} \varphi^{\mathrm{u}}_{s,i} + \sum_{i=1}^{n_{\mathrm{d}}} \varphi^{\mathrm{d}}_{s,i} \\ \begin{bmatrix} \varphi^{\mathrm{d}}_{s,i} & \mathrm{diag}(\zeta)^T_i \\ * & I \end{bmatrix} \succeq 0, \quad i = 1, 2, ..., n_{\mathrm{d}} \\ \begin{bmatrix} \varphi^{\mathrm{u}}_{s,i} & \mathrm{upper}(\zeta)^T_i \\ * & I \end{bmatrix} \succeq 0, \quad i = 1, 2, ..., n_{\mathrm{u}} \end{cases} \tag{47}$$

I. Proof of (35)

In (34b), the objective function is given by

$$L_s(\hat{\mathcal{Z}}^{(\kappa+1)}_s, \mathcal{Z}^{(\kappa)}_{\mathrm{a}}, \mathcal{Y}_s, \mathcal{A}^{(\kappa)}_s, \tilde{\mathcal{A}}^{(\kappa)}_s) = \underbrace{\frac{\tilde\rho}{2}\|\mathcal{Y}_s - \hat{\mathcal{Z}}^{(\kappa+1)}_s - \frac{1}{\tilde\rho}\tilde{\mathcal{A}}^{(\kappa)}_s\|^2_F}_{\mathcal{Y}_s \text{ involved}}$$
$$-\frac{1}{2\tilde\rho}\|\tilde{\mathcal{A}}^{(\kappa)}_s\| + \mathrm{Tr}\Big(\tilde{\mathcal{A}}^{(\kappa)T^*}_s \circ \hat{\mathcal{Z}}^{(\kappa+1)}_s\Big) + \frac{\rho}{2}\|\zeta^A_s(\hat{\mathcal{Z}}^{(\kappa+1)}_s) - \zeta^B_s(\mathcal{Z}^{(\kappa)}_{\mathrm{a}})\|^2_F$$
$$+\mathrm{Tr}\left(\mathcal{A}^{(\kappa)T^*}_s \circ \left(\zeta^A_s(\hat{\mathcal{Z}}^{(\kappa+1)}_s) - \zeta^B_s(\mathcal{Z}^{(\kappa)}_{\mathrm{a}})\right)\right) + \hat{J}_s(\hat{\mathcal{Z}}^{(\kappa+1)}_s)$$

and $\mathcal{Y}_s$ is involved only in the first term. Dropping other terms and $\tilde\rho/2$ in the first term results in (35).

J. The pseudocode of the above feasibility-embedded distributed approach: See Algorithm 1.

K. Computation realization and parameter settings

IPOPT interfaced by Pyomo, and MOSEK interfaced by CVXPY, are employed to sovle NLP and SDP in the case study, respectively. All computations are carried out on a Linux 64-Bit server with 2 Intel(R) Xeon(R) E5-2640 v4 @ 2.40GHz CPUs (a total of 40 processors provided), 125GB RAM. Distributed computing across multiple processors for line 5 of Algorithm 1 is realized using Ray.

Two generator settings are considered to simulate different operating modes or composition of generators. For the AU14Gen system, all generators are modelled as inverters in $\mathcal{N}_{v_{dm}}$, where both inertia and damping of all inverters are dispatchable. For other test power systems, generators are set in $\mathcal{N}_{v_o}$ (or $\mathcal{N}_g$), $\mathcal{N}_{v_d}$ (low inertia), $\mathcal{N}_{v_d}$ (high inertia), $\mathcal{N}_{v_m}$ and $\mathcal{N}_{v_{dm}}$ circularly. Parameters of each set of generators are given in Table IV. Transient reactances of synchronous generators are ignored for simplicity. The base MVA is 100 MVA. For each load, $d_{li} = 0.01$ p.u. $\cdot$ s/rad. For each test system, $\mathcal{D}$ contains four disturbances with parameters given in Table V. W_1 to W_5 are all set to identity matrices with proper dimension. $t_0 = 0$ s and $t_f = 30$ s. Frequency bounds $\overline{\omega}^k(t)$ and $\underline{\omega}^k(t)$ are give in Table VI, referring to draft NEM mainland frequency operating standards of interconnected systems [13]. $\overline{\delta} = 3\pi/4$.

Algorithm 1 Feasibility-embedded distributed approach

Input: N_S, Ξ_s, ρ, $\tilde\rho$, ϵ^{abs} and ϵ^{rel}
Output: $M\mathbb{1}, D\mathbb{1}$

1: Initialize $\mathcal{Z}^{(0)}_a$, $\mathcal{Y}^{(0)}_s$, $\mathcal{A}^{(0)}_s$, $\tilde{\mathcal{A}}^{(0)}_s$ and $\kappa \leftarrow -1$
2: **repeat**
3: $\quad\kappa \leftarrow \kappa + 1$
4: $\quad$**for** $s \leftarrow 1$ to N_S **do** $\hat{\mathcal{Z}}^{(\kappa+1)}_s \leftarrow$ Eq. (31) **end for**
5: $\quad\mathcal{Z}^{(\kappa+1)}_{\mathrm{a}} \leftarrow$ Eq. (23b)
6: $\quad$**for** $s \leftarrow 1$ to N_S **do**
7: $\quad\quad$**for each** $\mathcal{C}_{ej}$ in $\mathcal{K}_s$ **do**
8: $\quad\quad\quad\{\sigma^1_{\mathcal{C}_{ej}}, u^1_{\mathcal{C}_{ej}}, v^{1\,T}_{\mathcal{C}_{ej}}\} \leftarrow$ SVD for $\hat{\boldsymbol{Z}}^{(\kappa+1)}_{\mathcal{C}_{ej}} + \frac{1}{\tilde\rho}\hat{\boldsymbol{\Lambda}}^{(\kappa)}_{\mathcal{C}_{ej}}$
9: $\quad\quad$**end for**
10: $\quad\quad\mathcal{Y}^{(\kappa+1)}_s \leftarrow$ Eq. (36)
11: $\quad$**end for**
12: $\quad$**for** $s \leftarrow 1$ to N_S **do** $\{\mathcal{A}^{(\kappa+1)}_s, \tilde{\mathcal{A}}^{(\kappa+1)}_s\} \leftarrow$Eq. (23c, 37) **end for**
13: $\quad$Compute $r^{(\kappa+1)}$, $s^{(\kappa+1)}$, $\epsilon^{\mathrm{pri}(\kappa+1)}$ and $\epsilon^{\mathrm{dual}(\kappa+1)}$.
14: **until** $\|r^{(\kappa+1)}\|_2 < \epsilon^{\mathrm{pri}(\kappa+1)} \wedge \|s^{(\kappa+1)}\|_2 < \epsilon^{\mathrm{dual}(\kappa+1)}$
15: $m_j \leftarrow \boldsymbol{Z}^{\mathrm{md}}_{j,(1,2)}$, $d_j \leftarrow \boldsymbol{Z}^{\mathrm{md}}_{j,(1,3)}$, $\forall j \in \mathcal{N}_g$

In the NLP formulation of DID, 3rd-order Radua collocation and $n^k_t = 20$ are employed. In Approximation 1, $\theta_b = 0.580001$ to minimize the approximation error according to Appendix-E, where $\epsilon = 2.2155 \times 10^{-4}$. In the fill-reducing Cholesky factorization, $\beta_{cf} = 100$ can guarantee positive definiteness of $A_{adj} + \beta_{cf} I$ for all systems. In the feasibility-embedded distributed approach, $N_S = 40$, $\Xi_s = \{(k, 2s-1), (k, 2s)\}$ with $k \in \mathcal{D}$ and $s \in \mathbb{P}$. $\epsilon^{\mathrm{abs}} = 10^{-5}$ and $\epsilon^{\mathrm{rel}} = 10^{-3}$ [21].

TABLE IV
PARAMETERS OF GENERATORS.

Generator	$\underline{m}_i$	$\overline{m}_i$	$\overline{d}_i$	$\underline{d}_i$
$\mathcal{N}_{v_o}$ and $\mathcal{N}_g$	$0.5\widetilde{m}_i$	$0.5\widetilde{m}_i$	$0.5\widetilde{d}_i$	$0.5\widetilde{d}_i$
$\mathcal{N}_{v_d}$ (low inertia)	$0.01\widetilde{m}_i$	$0.01\overline{m}_i$	$0.01\widetilde{d}_i$	$\widetilde{d}_i$
$\mathcal{N}_{v_d}$ (high inertia)	$0.5\widetilde{m}_i$	$0.5\widetilde{m}_i$	$0.01\widetilde{d}_i$	$\widetilde{d}_i$
$\mathcal{N}_{v_m}$	$0.01\widetilde{m}_i$	$\widetilde{m}_i$	$0.5\widetilde{d}_i$	$0.5\widetilde{d}_i$
$\mathcal{N}_{v_{dm}}$	$0.01\widetilde{m}_i$	$\widetilde{m}_i$	$0.01\widetilde{d}_i$	$\widetilde{d}_i$

Note: $\widetilde{m}_i = \frac{10 p_{i,max}}{\omega_{syn}}$ with $p_{i,max}$ being maximal steady-state active power output of generator i and ω_{syn} being the synchronous angular speed. $\widetilde{d}_i = \frac{2p_{i,max}}{2\pi}$. $\overline{p}_g = -\underline{p}_g = 3p_{i,max}$ for each generator.

TABLE V
FREQUENCY BOUNDS.

Disturbance	Time interval	$\underline{\omega}^k(t)$	$\overline{\omega}^k(t)$
$\mathcal{D}_1$ and $\mathcal{D}_2$	{[0s,15s),[15s,30s]}	$\{49.5, 49.85\}$	{$50.5, 50.15$}
$\mathcal{D}_3$	$[0s, 30s]$	49.85	50.15
$\mathcal{D}_4$	{[$0s, 15s$),[15s,30s]}	$\{49, 49.5\}$	{$51, 50.5$}

TABLE VI
PARAMETERS OF DISTURBANCES.

Test system	$\mathcal{D}_1$	$\mathcal{D}_2$	$\mathcal{D}_3$	$\mathcal{D}_4$
AU14Gen	203	508	404	(212, 217)
IEEE 14-bus	2	9	6	(9, 14)
IEEE 39-bus	32	8	39	(17, 27)
IEEE 118-bus	25	54	89	(43, 44)
ACTIVSg200	127	100	155	(177, 58)

Note: The above table shows location of disturbances, with numbers denoting bus number. Test systems are at the equilibrium point at $t = t_0^-$, each disturbance occurs at $t = t_0$, and P_0 denotes the initial load power or generation power. For disturbances in $\mathcal{D}_1$, step amplitude is set to $-50\%P_0$, where P_0 denotes the initial load power for load buses or generation power for generator buses. For disturbances in $\mathcal{D}_2$, height of ramp and duration of ramp are set to $-50\%P_0$ and 5 s, respectively. Disturbances in $\mathcal{D}_3$ are emulated by a random power disturbance which changes its value randomly at a equal interval being 0.5 s according to a uniform distribution with the interval being $[-20\%P_0, +20\%P_0]$. Disturbances in $\mathcal{D}_4$ are assumed occurring at the middle of the branch, with short circuit resistance being 0, and being cleared by disconnecting the two sides breakers of the branch after 0.1 s. ρ_k is set to 0.15, 0.15, 0.6 and 0.1 for disturbances in $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and $\mathcal{D}_4$, respectively.