A Feedback-Based Regularized Primal-Dual Gradient Method for Time-Varying Nonconvex Optimization

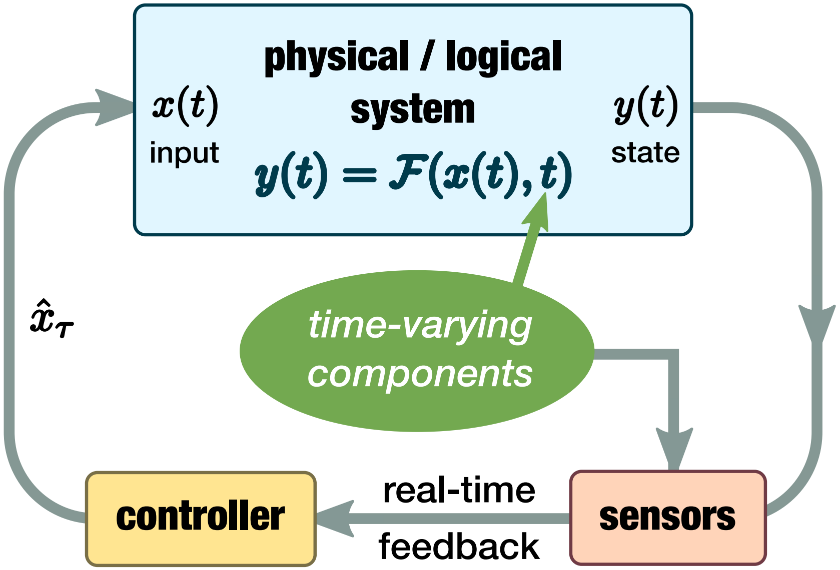
Yujie Tang Caltech

Emiliano Dall'Anese CU Boulder

Andrey Bernstein NREL

Steven H. Low Caltech

Optimal Operation of Time-Varying Systems

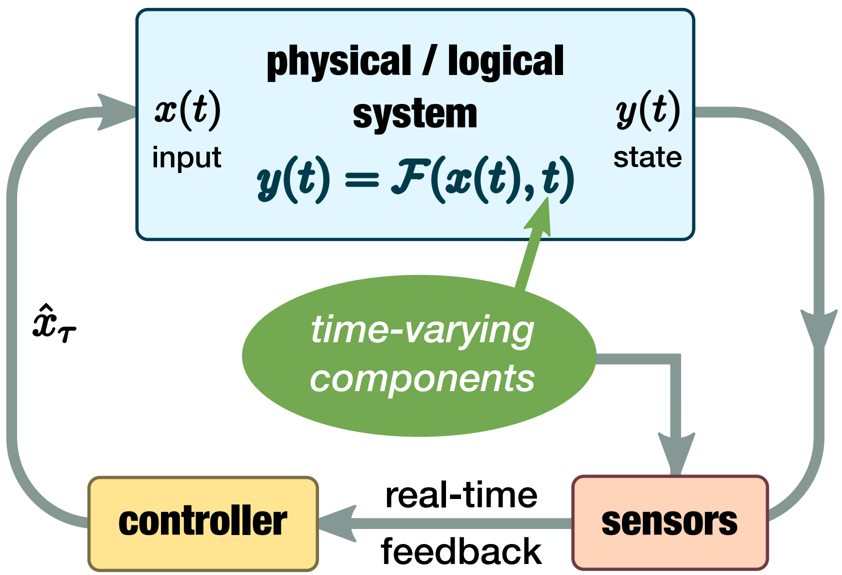


$$\min_{x \in \mathcal{X}(t)} \quad c(x, t)$$
$$\text{s.t.} \quad h(\mathcal{F}(x, t), t) \leq 0$$

$t \in [0, T]$

optimal trajectory $x^*(t)$

Optimal Operation of Time-Varying Systems



$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$

$$\tau = 1, 2, \ldots, \lfloor T/\Delta \rfloor$$

sampled optimal trajectory x_τ^*

Batch vs. Running

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \le 0$$

***batch* scheme**

for $\tau = 1, 2, \ldots, \lfloor T/\Delta \rfloor$

$\mathsf{P}_\tau \leftarrow$ problem data at $t = \tau\Delta$

repeat

$$z^k = T(z^{k-1}; \mathsf{P}_\tau)$$

until convergence

apply $x^*_\tau = \Pi(z^\infty)$

wait until $t = (\tau + 1)\Delta$

end for

should be finished within $(\tau\Delta, (\tau+1)\Delta)$

Batch vs. Running

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$

for $\tau = 1, 2, \ldots, \lfloor T/\Delta \rfloor$

$\mathsf{P}_\tau \leftarrow$ problem data at $t = \tau\Delta$

repeat

$z^k = T(z^{k-1}; \mathsf{P}_\tau)$

until convergence

apply $x_\tau^* = \Pi(z^\infty)$

wait until $t = (\tau + 1)\Delta$

end for

batch scheme

Works fine if

- computation is fast
- problem changes slowly so that sampling doesn't need to be fast

Batch vs. Running

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$

for $\tau = 1, 2, \ldots, \lfloor T/\Delta \rfloor$

$\mathsf{P}_\tau \leftarrow$ problem data at $t = \tau\Delta$

repeat

$$z^k = T(z^{k-1}; \mathsf{P}_\tau)$$

until convergence

apply $x_\tau^* = \Pi(z^\infty)$

wait until $t = (\tau + 1)\Delta$

end for

batch scheme

What if

- system is large and computation is slow?
- Δ needs to be small to capture fast varying costs and constraints?

Batch vs. Running

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$

for $\tau = 1, 2, \ldots, \lfloor T/\Delta \rfloor$

$\mathsf{P}_\tau \leftarrow$ problem data at $t = \tau\Delta$

repeat

$z^k = T(z^{k-1}; \mathsf{P}_\tau)$

until convergence

apply $x^*_\tau = \Pi(z^\infty)$

wait until $t = (\tau + 1)\Delta$

end for

for $\tau = 1, 2, \ldots, \lfloor T/\Delta \rfloor$

$\mathsf{P}_\tau \leftarrow$ problem data at $t = \tau\Delta$

$z_\tau = T(z_{\tau-1}; \mathsf{P}_\tau)$

apply $\hat{x}_\tau = \Pi(z_\tau)$

wait until $t = (\tau + 1)\Delta$

end for

running **scheme**

Batch vs. Running

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$

for $\tau = 1, 2, \ldots, \lfloor T/\Delta \rfloor$

$\mathsf{P}_\tau \leftarrow$ problem data at $t = \tau\Delta$

$z_\tau = T(z_{\tau-1}; \mathsf{P}_\tau)$

apply $\hat{x}_\tau = \Pi(z_\tau)$

wait until $t = (\tau + 1)\Delta$

end for

***running* scheme**

- problem data is updated *during* the iterations
- Δ can be further reduced to capture fast varying costs and constraints

Time-Varying Optimization

- Theory & algorithms
  - A. Y. Popkov. 2005.
  - Q. Ling and A. Ribeiro. 2014.
  - A. Simonetto and G. Leus. 2014.
  - C. Xi and U. A. Khan. 2016.
  - A. Simonetto. 2017.
  - A. Hauswirth, I. Subotic, S. Bolognani, G. Hug, and F. Dörfler. 2018.
  - A. Bernstein, E. Dall'Anese, and A. Simonetto. 2018.

Time-Varying Optimization

- Power system operation
  - E. Dall'Anese and A. Simonetto. 2016.
  - A. Bernstein and E. Dall'Anese. 2017.
  - A. Hauswirth, A. Zanardi, S. Bolognani, F. Dörfler, and G. Hug. 2017.
  - Y. Tang, K. Dvijotham, and S. Low. 2017.
- Wireless communication J. Chen and V. K. N. Lau. 2012.
- Sparse signal recovery A. Balavoine, C. J. Rozell, and J. Romberg. 2015.
- Social networks B. Baingana, P. Traganitis, G. Giannakis, and G. Mateos. 2016

- Online convex optimization – dynamic regret
  - A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan. 2015.
  - A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro. 2016.
  - T. Yang, L. Zhang, R. Jin, and J. Yi. 2016.

This Work

- Analysis of regularized primal-dual gradient method for time-varying nonconvex problems
  - nonconvexity: power networks
  - tracking performance
- Incorporating feedback measurement
- Application in power system operation

- Y. Tang, E. Dall'Anese, A. Bernstein and S. Low. Running primal-dual gradient method for time-varying nonconvex problems. arXiv prerpint arXiv:1812.00613, 2018.
- Y. Tang, E. Dall'Anese, A. Bernstein and S. H. Low. A feedback-based regularized primal-dual gradient method for time-varying nonconvex optimization. CDC2018.

Regularized Primal-Dual Gradient Method

$$\min_{x \in \mathcal{X}(t)} \quad c(x,t) \qquad \text{s.t.} \quad f(x,t) \leq 0$$

sample interval Δ →

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x) \qquad \text{s.t.} \quad f_\tau(x) \leq 0$$

regularized Lagrangian:

$$L_\tau^\epsilon(x,\lambda) = c_\tau(x) + \lambda^T f_\tau(x) - \frac{\epsilon}{2}\|\lambda\|^2$$

primal-dual gradient in the running scheme:

$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + J_{f_\tau}(\hat{x}_{\tau-1})^T \hat{\lambda}_{\tau-1}\right)\right]$$

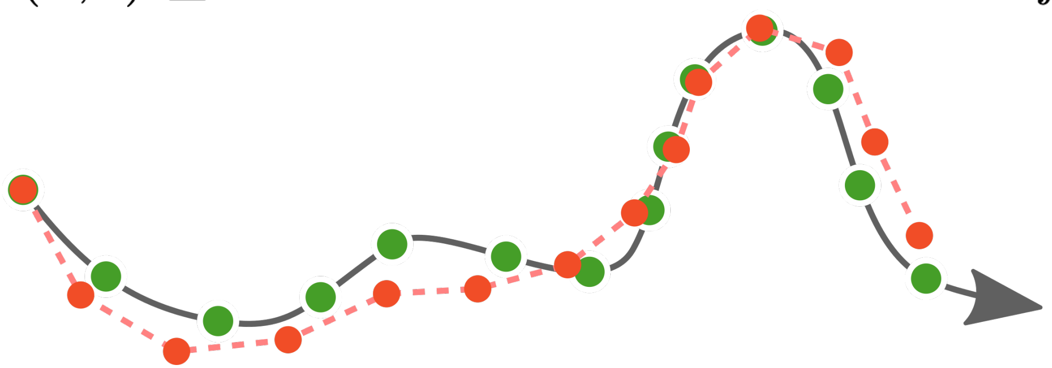
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}_+^m}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(f_\tau(\hat{x}_{\tau-1}) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

Tracking Performance

$$\min_{x \in \mathcal{X}(t)} \quad c(x,t)$$
$$\text{s.t.} \quad f(x,t) \leq 0$$

sample interval Δ

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad f_\tau(x) \leq 0$$



$$z^*(t) = \begin{bmatrix} x^*(t) \\ \lambda^*(t) \end{bmatrix}$$

a Lipschitz continuous KKT trajectory over $[0, T]$

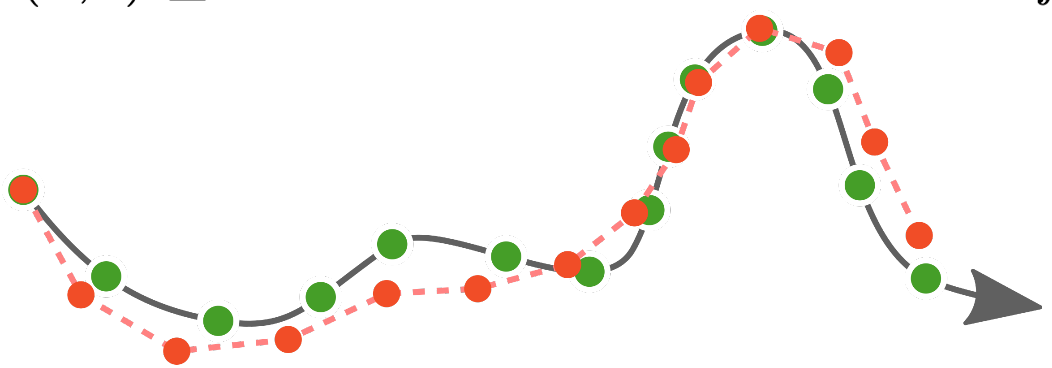
$$z^*_\tau = z^*(\tau\Delta)$$

Tracking Performance

$$\min_{x \in \mathcal{X}(t)} \quad c(x,t)$$
$$\text{s.t.} \quad f(x,t) \leq 0$$

sample interval Δ

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad f_\tau(x) \leq 0$$



tracking error

$$\|\hat{z}_\tau - z_\tau^*\|_\eta = \left(\|\hat{x}_\tau - x_\tau^*\|^2 + \eta^{-1}\|\hat{\lambda}_\tau - \lambda_\tau^*\|^2\right)^{1/2}$$

Tracking Performance

$$\min_{x\in\mathcal{X}(t)} \quad c(x,t) \qquad\qquad \min_{x\in\mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad f(x,t)\le 0 \qquad\qquad \text{s.t.} \quad f_\tau(x)\le 0$$

Theorem

Under certain conditions, suppose for some $\delta > 0$,

$$\Lambda_m(\delta) > M_{nc}(\delta) M_\lambda, \qquad M_\lambda := \sup_{t\in[0,T]} \|\lambda^*(t)\|$$

Then there exist sets of parameters such that the resulting sequence $\left(\hat{z}_\tau\right)_\tau$ satisfies

$$\|\hat{z}_\tau - z_\tau^*\|_\eta \le \frac{\rho}{1-\rho}\Delta \cdot \sup_{t\in[0,T]} \left\|\frac{d}{dt} z^*(t)\right\|_\eta + \frac{\sqrt{2\eta}\alpha\epsilon M_\lambda}{1-\rho}$$

when the initial point is sufficiently close to z_1^*.

$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + J_{f_\tau}(\hat{x}_{\tau-1})^T \hat{\lambda}_{\tau-1}\right)\right]$$
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}_+^m}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(f_\tau(\hat{x}_{\tau-1}) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

Tracking Performance

$$\min_{x \in \mathcal{X}(t)} \quad c(x,t) \qquad\qquad \min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad f(x,t) \le 0 \qquad\qquad \text{s.t.} \quad f_\tau(x) \le 0$$

Theorem

Under certain conditions, suppose

the problem is sufficiently convex in a neighborhood of $x^*(t)$

to overcome the nonlinearity of the nonconvex components

Then there exist sets of parameters such that the resulting sequence $\left(\hat{z}_\tau\right)_\tau$ satisfies

$$\|\hat{z}_\tau - z_\tau^*\|_\eta \le \frac{\rho}{1-\rho}\Delta \cdot \sup_{t\in[0,T]} \left\| \frac{d}{dt} z^*(t) \right\|_\eta + \frac{\sqrt{2\eta}\alpha\epsilon M_\lambda}{1-\rho}$$

$$M_\lambda := \sup_{t\in[0,T]} \|\lambda^*(t)\|$$

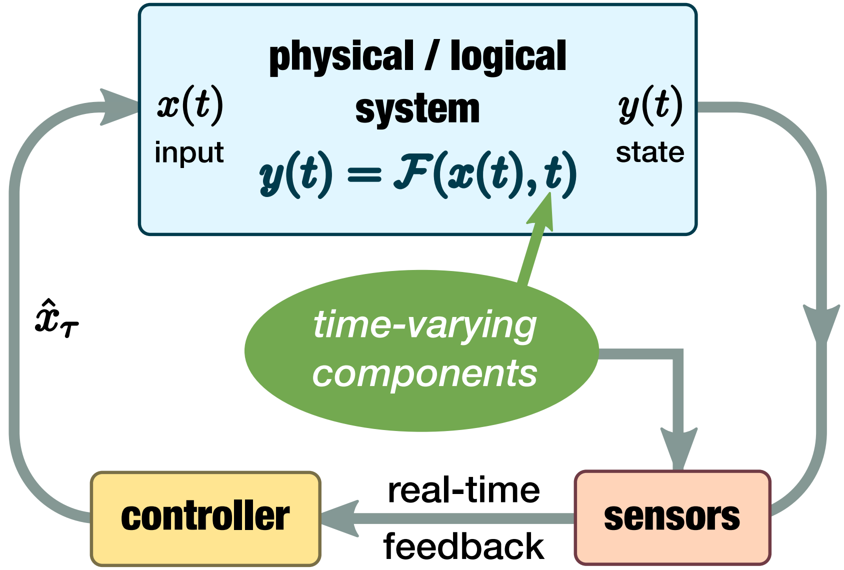
when the initial point is sufficiently close to z_1^*.

$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + J_{f_\tau}(\hat{x}_{\tau-1})^T \hat{\lambda}_{\tau-1}\right)\right]$$
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}_+^m}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(f_\tau(\hat{x}_{\tau-1}) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

Incorporating Feedback

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \le 0$$

$$h(y,t) = H(t)y + \beta(t)$$
$$h_\tau(y) = h(y, \tau\Delta)$$

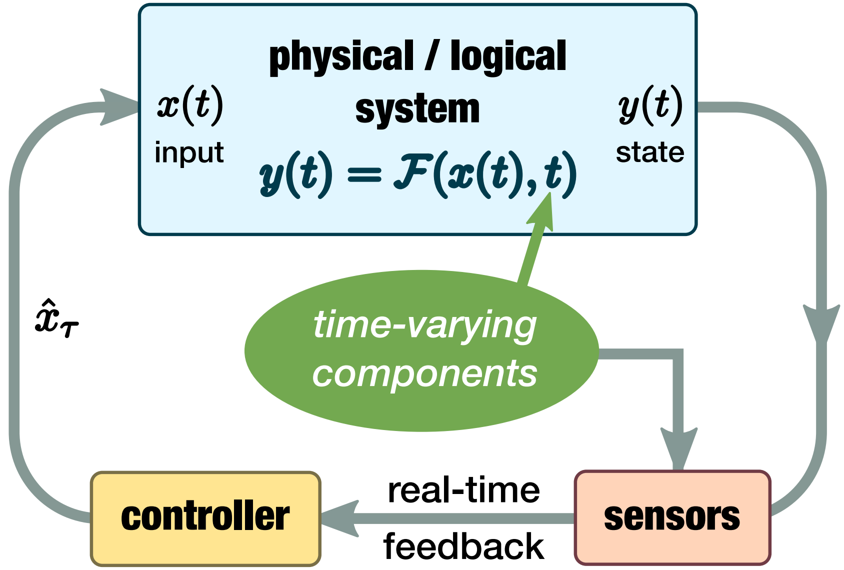


$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + [H_\tau J_{\mathcal{F}_\tau}(\hat{x}_{\tau-1})]^T \hat{\lambda}_{\tau-1}\right)\right]$$
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}^m_+}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(h_\tau(\mathcal{F}_\tau(\hat{x}_{\tau-1})) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

Can be replaced by measurement data

Incorporating Feedback

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$



$$h(y, t) = H(t)y + \beta(t)$$
$$h_\tau(y) = h(y, \tau\Delta)$$

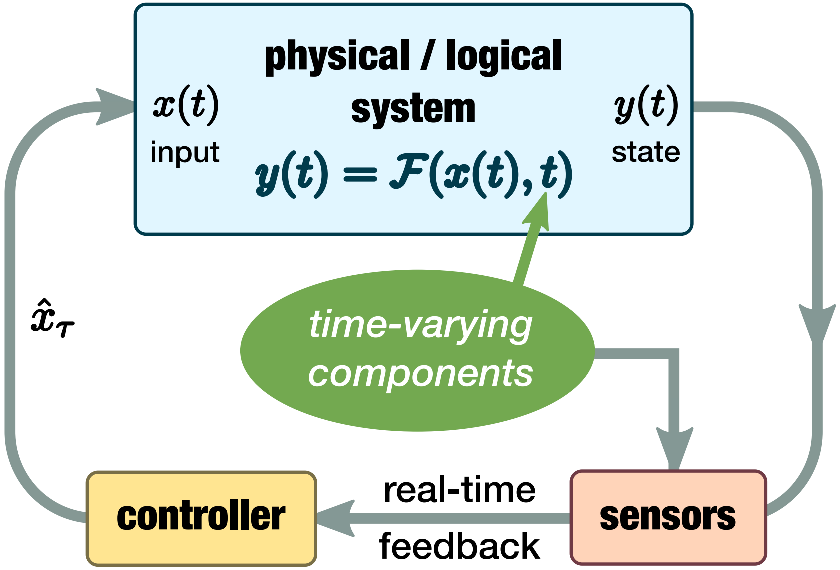
$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + [H_\tau J_{\mathcal{F}_\tau}(\hat{x}_{\tau-1})]^T \hat{\lambda}_{\tau-1}\right)\right]$$
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}^m_+}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(h_\tau(\check{y}_\tau) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

$\check{y}_\tau =$ measured valued of $\mathcal{F}_\tau(\hat{x}_{\tau-1})$

Incorporating Feedback

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$

$$h(y,t) = H(t)y + \beta(t)$$
$$h_\tau(y) = h(y, \tau\Delta)$$



$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + [H_\tau J_{\mathcal{F}_\tau}(\hat{x}_{\tau-1})]^T \hat{\lambda}_{\tau-1}\right)\right]$$
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}^m_+}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(h_\tau(\check{y}_\tau) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

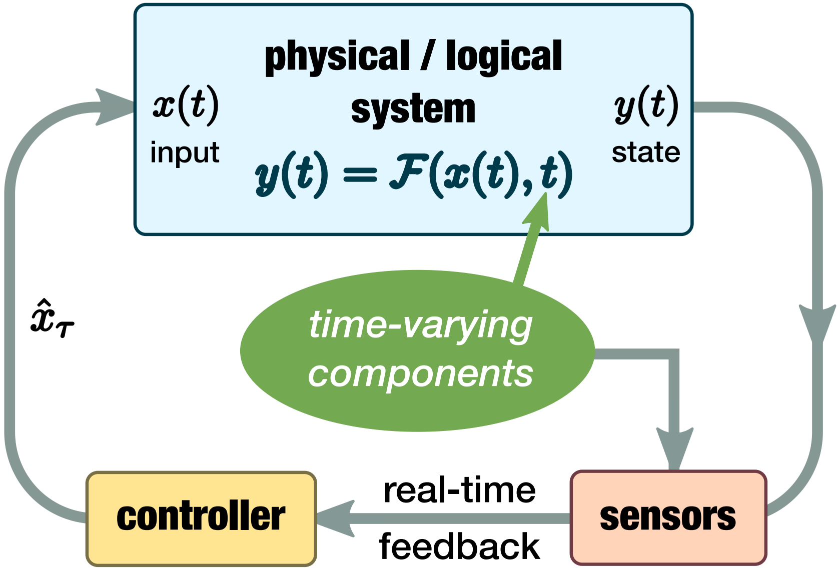
There exist a continuous $\mathsf{J}_\tau(x,y)$ and a nondecreasing $e_J(\delta)$

s.t. $\|\mathsf{J}_\tau(x, \mathcal{F}_\tau(x)) - J_{\mathcal{F}_\tau}(x)\| \leq e_J(\delta), \quad \forall \|x - x^*_\tau\| \leq \delta$

Incorporating Feedback

$$\min_{x \in \mathcal{X}_\tau} \quad c_\tau(x)$$
$$\text{s.t.} \quad h_\tau(\mathcal{F}_\tau(x)) \leq 0$$

$$h(y,t) = H(t)y + \beta(t)$$
$$h_\tau(y) = h(y, \tau\Delta)$$



$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + [H_\tau \mathsf{J}_\tau(\hat{x}_{\tau-1}, \check{y}_\tau)]^T \hat{\lambda}_{\tau-1}\right)\right]$$
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}^m_+}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(h_\tau(\check{y}_\tau) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

Tracking Performance

$$\begin{aligned} \min_{x \in \mathcal{X}(t)} \quad & c(x,t) \\ \text{s.t.} \quad & h(\mathcal{F}(x,t),t) \le 0 \end{aligned} \qquad \begin{aligned} \min_{x \in \mathcal{X}_\tau} \quad & c_\tau(x) \\ \text{s.t.} \quad & h_\tau(\mathcal{F}_\tau(x)) \le 0 \end{aligned}$$

Theorem

Under certain conditions, suppose

$$\|\check{y}_\tau - \mathcal{F}_\tau(\hat{x}_{\tau-1})\| \le e_y \qquad L_h := \sup_{t \in [0,T]} \|H(t)\|$$

Then the sequence $\left(\hat{z}_\tau\right)_\tau$ satisfies

$$\|\hat{z}_\tau - z_\tau^*\|_\eta \le \frac{\rho}{1 - \rho - \alpha\sqrt{\eta} e_J} \Delta \cdot \sup_{t \in [0,T]} \left\| \frac{d}{dt} z^*(t) \right\|_\eta + \frac{\sqrt{2\eta}\alpha\epsilon(M_\lambda + \epsilon^{-1} L_h e_y) + \alpha\sqrt{\eta} e_J M_\lambda}{1 - \rho - \alpha\sqrt{\eta} e_J}$$

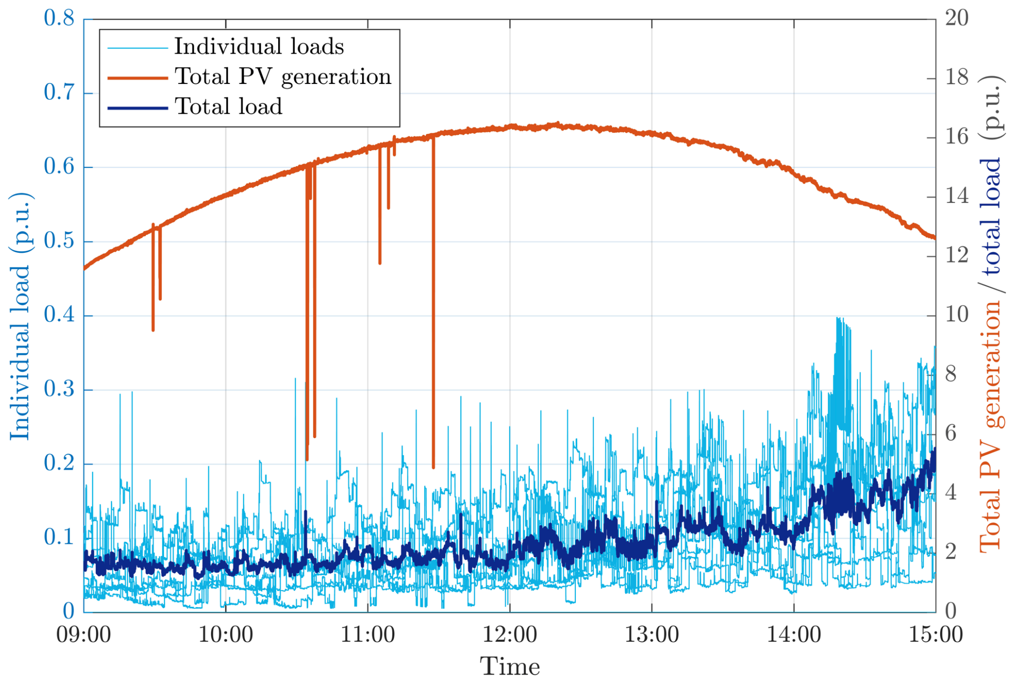
when the initial point is sufficiently close to z_1^*.

$$\hat{x}_\tau = \mathcal{P}_{\mathcal{X}_\tau}\left[\hat{x}_{\tau-1} - \alpha\left(\nabla c_\tau(\hat{x}_{\tau-1}) + [H_\tau \mathsf{J}_\tau(\hat{x}_{\tau-1}, \check{y}_\tau)]^T \hat{\lambda}_{\tau-1}\right)\right]$$

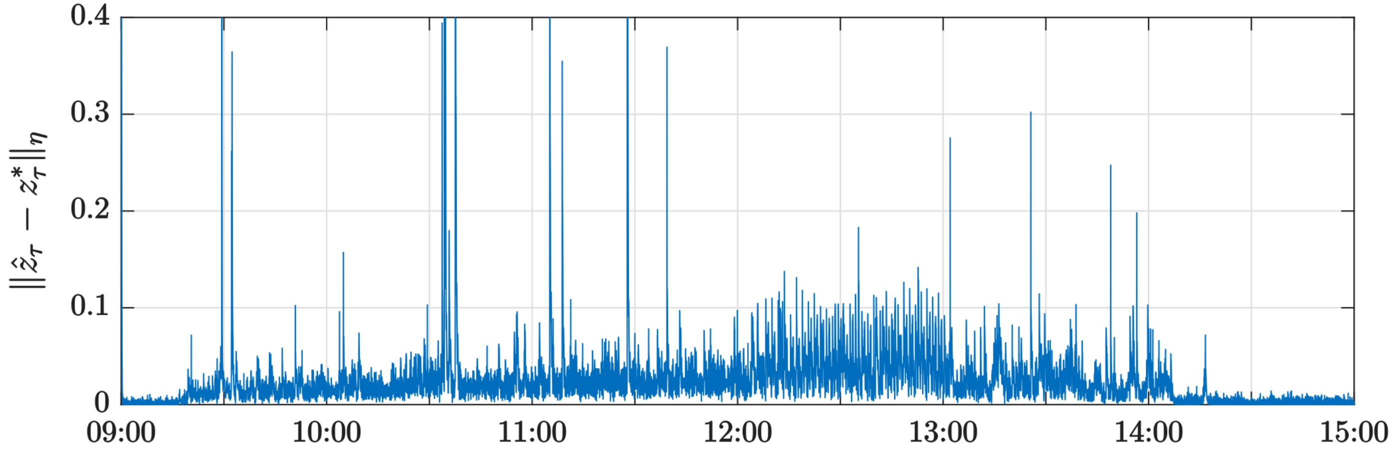
$$\hat{\lambda}_\tau = \mathcal{P}_{\mathbb{R}_+^m}\left[\hat{\lambda}_{\tau-1} + \eta\alpha\left(h_\tau(\check{y}_\tau) - \epsilon\hat{\lambda}_{\tau-1}\right)\right]$$

Numerical Example

- Power system test case.
- Single-phase version of the IEEE 37 node test feeder
- High penetration of PV systems



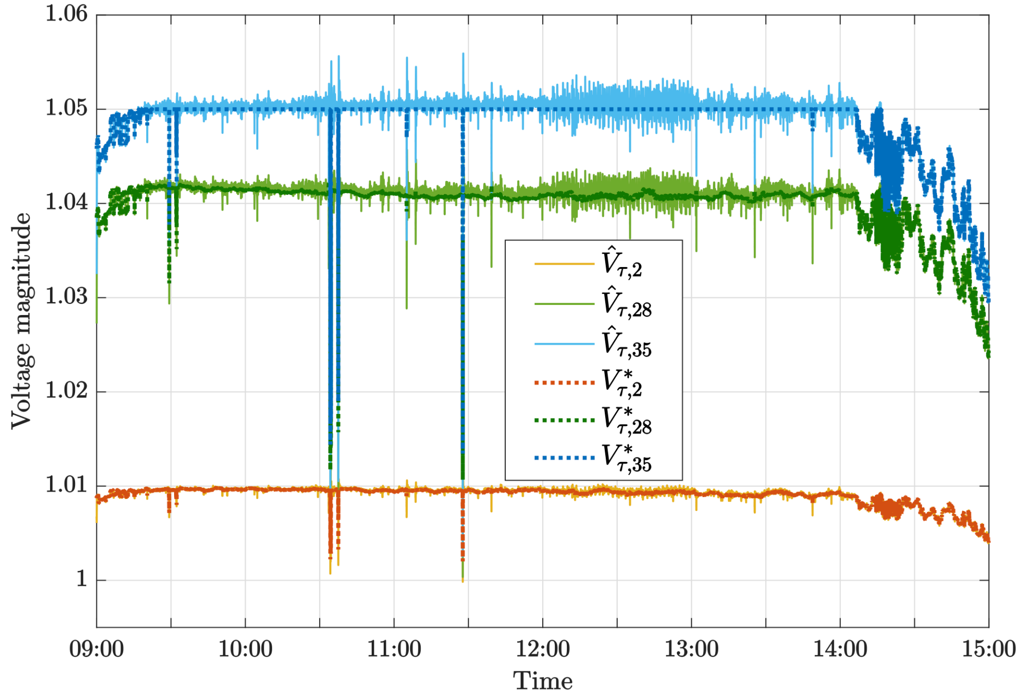
Numerical Example



$$\frac{1}{T}\sum_\tau \|\hat{z}_\tau - z^*_\tau\|_\eta = 2.56 \times 10^{-2}$$

$$\frac{1}{T}\sum_\tau \frac{\|\hat{z}_\tau - z^*_\tau\|_\eta}{\|z^*_\tau\|_\eta} = 7.02 \times 10^{-3}$$

Numerical Example



$$\frac{1}{T}\sum_{\tau,j}\left(\left[\hat{V}_{\tau,j}-V_{\max}\right]_+ + \left[V_{\min}-\hat{V}_{\tau,j}\right]_+\right) = 3.67\times 10^{-4}$$

Future Directions

- Different metric for tracking performance
- Distributed algorithm
- Jumps in the optimal trajectory
- Coupling in the time domain

For more details

- Y. Tang, E. Dall'Anese, A. Bernstein and S. Low. Running primal-dual gradient method for time-varying nonconvex problems. arXiv prerpint arXiv:1812.00613, 2018.
- Y. Tang, E. Dall'Anese, A. Bernstein and S. H. Low. A feedback-based regularized primal-dual gradient method for time-varying nonconvex optimization. CDC2018.