Distributed optimization for multiagent systems

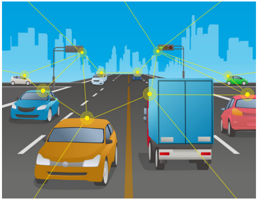
Ashish Cherukuri
Postdoctoral Researcher
Automatic Control Laboratory
ETH Zürich, Switzerland
Email: `cashish@control.ee.ethz.ch`
Website: `people.ee.ethz.ch/∼cashish/`

Indian Institute of Technology, Bombay
August 28, 2018

Joint work with: Bahman Gharesifard (Queen's University)
Enrique Mallada (JHU)
Steven Low (Caltech)
Jorge Cortés (UCSD)

Networked cyber-physical systems (CPS)

Energy network



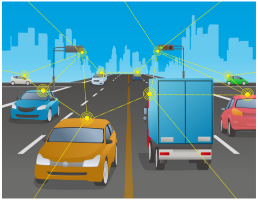
Transportation network

Supply-chain network

Smart city

Networked cyber-physical systems (CPS)

Energy network



Transportation network

Supply-chain network

Smart city

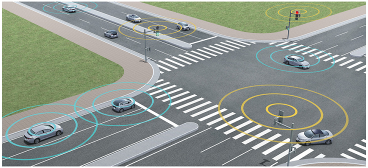
To achieve reliable, robust, secure, and efficient performance

Networked CPS: optimization challenges

Objective: reach optimizers
Path: algorithms with desirable properties

Problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{F} \end{aligned}$$



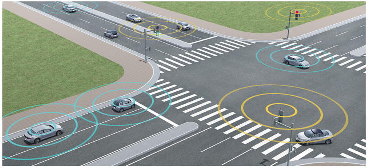
Networked CPS: optimization challenges

Objective: reach optimizers
Path: algorithms with desirable properties
Challenges:

- size
- time-scales
- perturbations
- uncertainty
- privacy & security

Problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{F} \end{aligned}$$



Networked CPS: optimization challenges

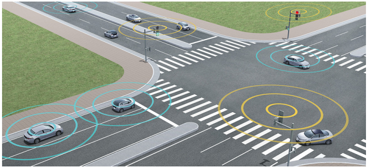
Objective: reach optimizers
Path: algorithms with desirable properties
Challenges:

- size
- time-scales
- perturbations
- uncertainty
- privacy & security

Problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{F} \end{aligned}$$

Distributed algorithms:



Networked CPS: optimization challenges

Objective: reach optimizers
Path: algorithms with desirable properties
Challenges:

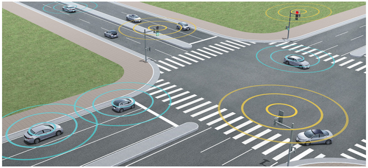
- size
- time-scales
- perturbations
- uncertainty
- privacy & security

Problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{F} \end{aligned}$$

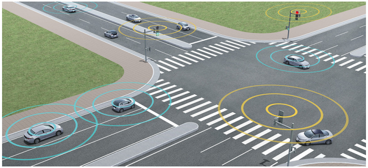
Distributed algorithms:

- continuous-time stability analysis
- optimization theory
- algebraic graph theory



Networked CPS: challenges

Game theory



Game: strategic scenario

- players: $1, \dots, n$
- actions: x_i for player i
- utility: $u_i(x_1, \dots, x_n)$ for i

Players maximize their utility
Equilibrium $x_1^{\mathrm{eq}}, \dots, x_n^{\mathrm{eq}}$

Social welfare

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{F} \end{aligned}$$

Optimizer $x_1^*, \dots, x_n^*$

Networked CPS: challenges

Game theory

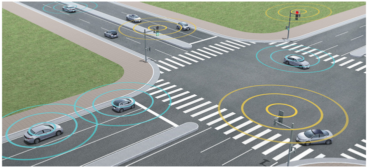
Objective: reach efficient equilibria
Path: utilities with desirable properties
Challenges:

- predicting x^{eq}
- changing utilities for $x^{\mathrm{eq}} = x^*$
- all previous ones

Game: strategic scenario

- players: $1, \dots, n$
- actions: x_i for player i
- utility: $u_i(x_1, \dots, x_n)$ for i

Players maximize their utility
Equilibrium $x_1^{\mathrm{eq}}, \dots, x_n^{\mathrm{eq}}$



Social welfare

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{F} \end{aligned}$$

Optimizer $x_1^*, \dots, x_n^*$

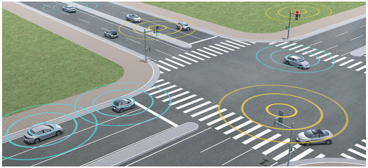
Networked CPS: challenges

Game theory

Objective: reach efficient equilibria
Path: utilities with desirable properties
Challenges:

- predicting x^{eq}
- changing utilities for $x^{\text{eq}} = x^*$
- all previous ones

Dynamic analysis of competition:
behaviour around Nash equilibrium



Game: strategic scenario

- players: $1, \ldots, n$
- actions: x_i for player i
- utility: $u_i(x_1, \ldots, x_n)$ for i

Players maximize their utility
Equilibrium $x_1^{\text{eq}}, \ldots, x_n^{\text{eq}}$

Social welfare

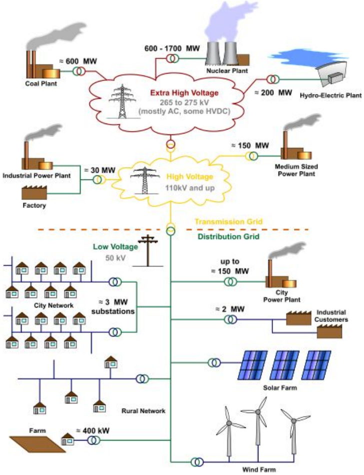
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{F} \end{aligned}$$

Optimizer $x_1^*, \ldots, x_n^*$

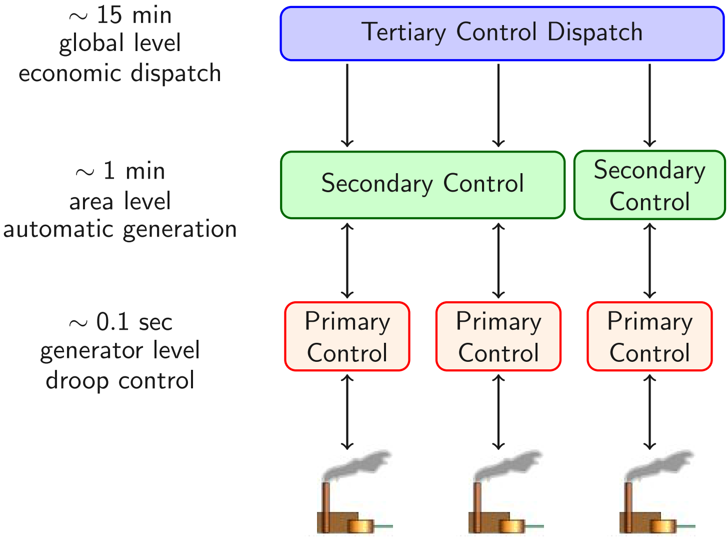
Electrical power network

Objectives:

- balance load and generation
- restore nominal frequency
  - guarantee cost efficiency
  - satisfy physical constraints
  - ensure security & reliability



Traditional approach: hierarchy of controllers



Tertiary control/dispatch: future challenges

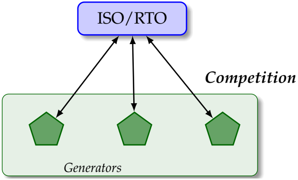
Current practice:

- generators submit (closed) bids to the ISO
- ISO solves the following problem

Security constrained OPF

$$\begin{aligned} \underset{P}{\text{minimize}} \quad & \text{payment}(P) \\ \text{subject to} \quad & P \in \mathcal{F} \end{aligned}$$

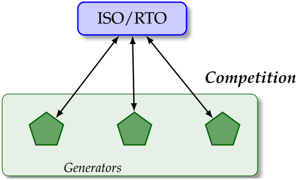
- ISO sends P_i to each generator i



Tertiary control/dispatch: future challenges

Current practice:

- generators submit (closed) bids to the ISO
- ISO solves the following problem



Security constrained OPF

$$\begin{aligned} \underset{P}{\text{minimize}} \quad & \text{payment}(P) \\ \text{subject to} \quad & P \in \mathcal{F} \end{aligned}$$

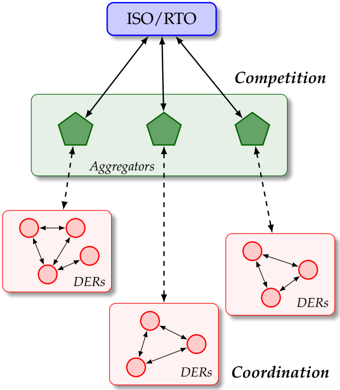
- ISO sends P_i to each generator i

Future challenge: Too many generators; shorter time-scales

- how to integrate them into the existing system?
- can we avoid market manipulation, congestion, failures?

Coordination and competition in dispatch

- at the top-level, aggregators compete and at the bottom-level, DERs coordinate



CAISO. "Expanded metering and telemetry options phase 2 - distributed energy resource provider", 2015.

Coordination in Dispatch

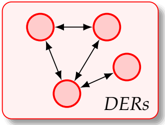
Coordinating the DERs

Economic Dispatch (ED) Problem

$$\min_{P} \quad f(P) := \sum_{i=1}^{n} f_i(P_i)$$

$$\text{s.t} \quad \sum_{i=1}^{n} P_i = \mathbf{1}_n^\top P = \ell$$

$$P_i^m \leq P_i \leq P_i^M, \quad \text{for all } i$$



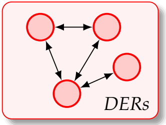
Coordination

$$\mathbf{1}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Coordinating the DERs

Economic Dispatch (ED) Problem

$$\begin{aligned} \min_{P} \quad & f(P) \\ \text{s.t} \quad & \mathbf{1}_n^\top P = \ell \\ & P_i^m \leq P_i \leq P_i^M, \ \text{for all } i \end{aligned}$$

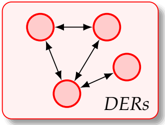


Coordination

Coordinating the DERs

Economic Dispatch (ED) Problem

$$\begin{aligned} \min_{P} \quad & f(P) \\ \text{s.t} \quad & \mathbf{1}_n^\top P = \ell \\ & P_i^m \le P_i \le P_i^M, \quad \text{for all } i \end{aligned}$$



Coordination

Objective: design distributed algorithm

- solves the ED problem globally
- able to handle time-varying loads
- handle plug-n-play

Communication network

- connected network
- gen i knows f_i; controls P_i
- gen i comm. with neighbors

Laplacian-gradient **dynamics**

ED Problem

$$\begin{aligned} \min \quad & f(P) \\ \text{s.t} \quad & \mathbf{1}_n^\top P = \ell \end{aligned}$$

KKT conditions: $\nu^* \mathbf{1}_n = \nabla f(P^*)$ and $\mathbf{1}_n^\top P^* = \ell$

Consensus dynamics: $\dot{x} = -\mathrm{L}x$ leads $x(t) \to \nu \mathbf{1}_n$
where L is the Laplacian matrix

Laplacian-gradient **dynamics**

ED Problem

$$\begin{aligned} \min \quad & f(P) \\ \text{s.t} \quad & \mathbf{1}_n^\top P = \ell \end{aligned}$$

Laplacian-gradient dynamics

$$\dot{P} = -\mathsf{L}\nabla f(P)$$

where

$$\nabla f(P)^\top = [\nabla f_1(P_1), \ldots, \nabla f_n(P_n)]$$

(L is p.s.d $n \times n$ matrix with $\mathbf{1}_n^\top \mathsf{L} = \mathsf{L}\mathbf{1}_n = 0$)

Laplacian-gradient **dynamics**

Laplacian-gradient dynamics

$$\dot{P} = -\mathsf{L}\nabla f(P)$$

ED Problem

$$\begin{aligned} \min \quad & f(P) \\ \text{s.t} \quad & \mathbf{1}_n^\top P = \ell \end{aligned}$$

where

$$\nabla f(P)^\top = [\nabla f_1(P_1), \ldots, \nabla f_n(P_n)]$$

(L is p.s.d $n \times n$ matrix with $\mathbf{1}_n^\top \mathsf{L} = \mathsf{L}\mathbf{1}_n = 0$)

Discrete-time algorithm:

- iteration: $P(k+1) = \mathtt{Alg}_k(P(k))$
- trajectory: $P(1), P(2), \ldots$ & $P(k) \to P^*$

Continuous-time algorithm:

- iteration: $\dot{P} = X_{\mathtt{Alg}}(P)$
- evolution: $t \mapsto P(t)$ & $P(t) \to P^*$

Laplacian-gradient **dynamics**

ED Problem

$$\begin{aligned} \min \quad & f(P) \\ \text{s.t} \quad & \mathbf{1}_n^\top P = \ell \end{aligned}$$

Laplacian-gradient dynamics

$$\dot{P} = -\mathrm{L}\nabla f(P)$$

where

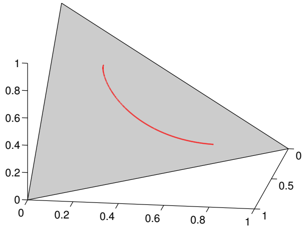
$$\nabla f(P)^\top = [\nabla f_1(P_1), \ldots, \nabla f_n(P_n)]$$

(L is p.s.d $n \times n$ matrix with $\mathbf{1}_n^\top \mathrm{L} = \mathrm{L}\mathbf{1}_n = 0$)

- distributed implementation: $\dot{P}_i = -\sum_{j\in\mathcal{N}_i} a_{ij}(\nabla f_i(P_i) - \nabla f_j(P_j))$
- load condition conserved: $\frac{d}{dt}(\mathbf{1}_n^\top P) = -\mathbf{1}_n^\top \mathrm{L}\nabla f(P) = 0$
- f nonincreasing: $\frac{d}{dt} f(P(t)) = -\nabla f(P)^\top \mathrm{L}\nabla f(P) \leq 0$

Centralized global solution

$$\dot{P} = -\mathsf{L}\nabla f(P) + \frac{1}{n}(\ell - \mathbf{1}_n^\top P)\mathbf{1}_n$$



- mismatch dynamics: $\frac{d}{dt}(\ell - \mathbf{1}_n^\top P) = -(\ell - \mathbf{1}_n^\top P)$
- on load satisfaction, it reduces to `Laplacian-gradient` dyn
- conv. analysis using refined LaSalle Invaraince (Arsie and Ebenbauer '10)
  - $V_1(P) = (\ell - \mathbf{1}_n^\top P)^2$
  - $V_2(P) = f(P)$

How to get a distributed solution?

Laplacian-gradient dyn.

$$\dot{P}_i = (-\mathsf{L}\nabla f(P))_i + \nu_1 z_i$$

Generation levels $P_1, \ldots, P_n$ $\Downarrow$

$\Uparrow$ Load mismatch estimate $z_1, z_2, \ldots, z_n$

dynamic average consensus (dac)

$$\begin{bmatrix} \dot{z} \\ \dot{v} \end{bmatrix} = A(\alpha, \beta) \begin{bmatrix} z \\ v \end{bmatrix} + \nu_2 \begin{bmatrix} \ell e_r - P \\ 0 \end{bmatrix}$$

- Each unit i has estimator $z_i \in \mathbb{R}$ tracking average signal $t \mapsto \frac{1}{n}(\ell - \mathbf{1}_n^\top P(t))$

Interconnected systems

- bottom component estimates evolving load mismatch given generation
- top component adjusts generation levels based on optimization of objective & estimate of load mismatch

Theoretical guarantees of $\mathtt{L}\nabla+\mathtt{dac}$ dynamics

Theorem (Convergence of $\mathtt{L}\nabla+\mathtt{dac}$ dynamics)

For $\alpha, \beta, \nu_1, \nu_2 > 0$ satisfying an inequality:

1. *the P-component of trajectories of $\mathtt{L}\nabla+\mathtt{dac}$ dynamics starting with $\mathbf{1}_n^\top v = 0$ converge to a solution of the ED problem*
2. *load-mismatch dynamics is exponentially stable*

[A. Cherukuri & J. Cortés, Automatica, 2016]

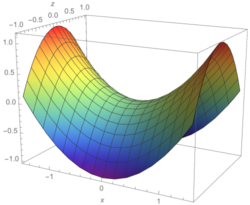
Performance guarantees ($\mathtt{L}\nabla+\mathtt{dac}$ dynamics)

- global convergence
- load mismatch dynamics is ISS
- dynamic loads tracked with ultimate bound
- robust to intermittent generation

Network optimization

Convex optimization

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t} \quad & g(x) \leq 0 \\
& Ax = b
\end{aligned}
$$

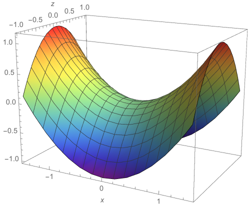


- Lagrangian: $L(x, y, z) = f(x) + y^\top g(x) + z^\top (Ax - b)$
- Primal-dual optimizers $\Leftrightarrow$ saddle points of L (over $\mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m$)
  - $L(x_*, y, z) \leq L(x_*, y_*, z_*) \leq L(x, y_*, z_*)$ for all x, z and $y \geq 0$

Network optimization

Convex optimization

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & g(x) \le 0 \\ & Ax = b \end{aligned}$$



- Lagrangian: $L(x, y, z) = f(x) + y^\top g(x) + z^\top (Ax - b)$
- Primal-dual optimizers $\Leftrightarrow$ saddle points of L (over $\mathbb{R}^n \times \mathbb{R}^p_{\ge 0} \times \mathbb{R}^m$)
  - $L(x_*, y, z) \le L(x_*, y_*, z_*) \le L(x, y_*, z_*)$ for all x, z and $y \ge 0$

Saddle-point dynamics

$$\begin{aligned} \dot{x} &= -\nabla_x L(x, y, z) \\ \dot{y} &= [\nabla_y L(x, y, z)]^+_y \\ \dot{z} &= \nabla_z L(x, y, z) \end{aligned}$$

$$[a]^+_b = \begin{cases} a & \text{if } a \ge 0 \text{ or } b > 0 \\ 0 & \text{otherwise} \end{cases}$$

Network optimization

Convex optimization

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & g(x) \leq 0 \\ & Ax = b \end{aligned}$$

- additive cost: $f(x) = \sum_{i=1}^{n} f_i(x_i)$
- local constraints:
  - g_k depends on some x_i and $\{x_j\}_{j \in \mathcal{N}_i}$
  - $(Ax)_k$ depends on some x_i and $\{x_j\}_{j \in \mathcal{N}_i}$

- Lagrangian: $L(x, y, z) = f(x) + y^\top g(x) + z^\top (Ax - b)$
- Primal-dual optimizers $\Leftrightarrow$ saddle points of L (over $\mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m$)
  - $L(x_*, y, z) \leq L(x_*, y_*, z_*) \leq L(x, y_*, z_*)$ for all x, z and $y \geq 0$

Saddle-point dynamics

$$\begin{aligned} \dot{x} &= -\nabla_x L(x, y, z) \\ \dot{y} &= [\nabla_y L(x, y, z)]_y^+ \\ \dot{z} &= \nabla_z L(x, y, z) \end{aligned}$$

$$[a]_b^+ = \begin{cases} a & \text{if } a \geq 0 \text{ or } b > 0 \\ 0 & \text{otherwise} \end{cases}$$

This dynamics is distributed for additive cost and local constraints!
When does this dynamics converge?

Network optimization

Convex optimization

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & g(x) \leq 0 \\ & Ax = b \end{aligned}$$

- additive cost: $f(x) = \sum_{i=1}^{n} f_i(x_i)$
- local constraints:
  - g_k depends on some x_i and $\{x_j\}_{j \in \mathcal{N}_i}$
  - $(Ax)_k$ depends on some x_i and $\{x_j\}_{j \in \mathcal{N}_i}$

- Lagrangian: $L(x, y, z) = f(x) + y^\top g(x) + z^\top (Ax - b)$
- Primal-dual optimizers $\Leftrightarrow$ saddle points of L (over $\mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m$)
  - $L(x_*, y, z) \leq L(x_*, y_*, z_*) \leq L(x, y_*, z_*)$ for all x, z and $y \geq 0$

Saddle-point dynamics

$$\begin{aligned} \dot{x} &= -\nabla_x L(x, y, z) \\ \dot{y} &= [\nabla_y L(x, y, z)]_y^+ \\ \dot{z} &= \nabla_z L(x, y, z) \end{aligned}$$

$$[a]_b^+ = \begin{cases} a & \text{if } a \geq 0 \text{ or } b > 0 \\ 0 & \text{otherwise} \end{cases}$$

[A. Cherukuri & B. Gharesifard & J. Cortés, SICON, 2017]
[A. Cherukuri & E. Mallada & J. Cortés, SCL, 2016]
[A. Cherukuri & E. Mallada & S. Low & J. Cortés, TAC, 2018]

Data-driven distributed optimization

Problem statement

Stochastic Optimization

$$\inf_{x \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}}[f(x, \xi)]$$

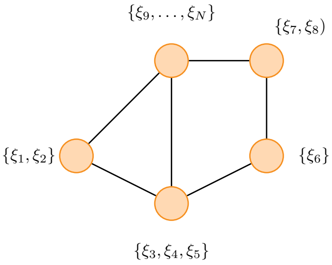
- $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, $(x, \xi) \mapsto f(x, \xi)$
    - continuously differentiable
    - convex-concave in (x, ξ)
- uncertainty ξ with prob. dist. $\mathbb{P}$ (unknown)

Problem statement

Stochastic Optimization

$$\inf_{x \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}}[f(x,\xi)]$$

- $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, $(x,\xi) \mapsto f(x,\xi)$
  - continuously differentiable
  - convex-concave in (x,ξ)
- uncertainty ξ with prob. dist. $\mathbb{P}$ (unknown)



Multiagent setup:

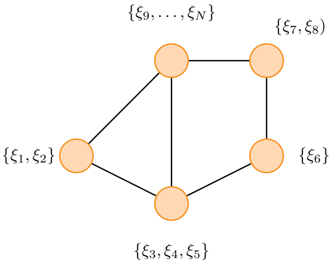
- n agents, communicating via an undirected graph $(\mathcal{V}, \mathcal{E})$
- each agent gathers i.i.d samples collected in $\widehat{\Xi}_i$, $\widehat{\Xi}_i \cap \widehat{\Xi}_j = \emptyset$
- total data $\widehat{\Xi} = \cup_{i=1}^n \widehat{\Xi}_i = \{\widehat{\xi}^k\}_{k=1}^N$

Problem statement

Stochastic Optimization

$$\inf_{x \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}}[f(x,\xi)]$$

- $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, $(x,\xi) \mapsto f(x,\xi)$
  - continuously differentiable
  - convex-concave in (x,ξ)
- uncertainty ξ with prob. dist. $\mathbb{P}$ (unknown)



Multiagent setup:

- n agents, communicating via an undirected graph $(\mathcal{V}, \mathcal{E})$
- each agent gathers i.i.d samples collected in $\widehat{\Xi}_i$, $\widehat{\Xi}_i \cap \widehat{\Xi}_j = \emptyset$
- total data $\widehat{\Xi} = \cup_{i=1}^n \widehat{\Xi}_i = \{\widehat{\xi}^k\}_{k=1}^N$

Goal for agents: find, in a distributed manner, approximate optimizer $\widehat{x}_N \in \mathbb{R}^d$ having guaranteed performance bounds

Background: Data-driven stochastic optimization

- $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ cont. diff.
- $\xi \sim \mathbb{P}$ (unknown)
- N i.i.d samples $\widehat{\Xi} := \{\widehat{\xi}^k\}_{k=1}^N$ are given
- $\widehat{\Xi}$ is a r.v.; support $(\mathbb{R}^m)^N$ and dist. $\mathbb{P}^N$

Stochastic Optimization

$$\inf_{x \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}}[f(x, \xi)]$$

Background: Data-driven stochastic optimization

- $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ cont. diff.
- $\xi \sim \mathbb{P}$ (unknown)
- N i.i.d samples $\widehat{\Xi} := \{\widehat{\xi}^k\}_{k=1}^N$ are given
- $\widehat{\Xi}$ is a r.v.; support $(\mathbb{R}^m)^N$ and dist. $\mathbb{P}^N$

Stochastic Optimization

$$\inf_{x \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}}[f(x, \xi)]$$

Goal: find a (data-driven) solution $\widehat{x}_N$ having:

- *finite-sample guarantee:*

$$\mathbb{P}^N\Big(\mathbb{E}_{\mathbb{P}}[f(\widehat{x}_N, \xi)] \leq \widehat{J}_N\Big) \geq 1 - \beta$$

  $\widehat{J}_N$ is the certificate and $1 - \beta$ is the reliability ($\beta \in (0, 1)$)
- *tractability:* solving for $\widehat{x}_N$ is a convex program

Background: Ambiguity sets and DRO

Approach:

- find an *ambiguity set* $\widehat{\mathcal{P}}_N$ of prob. dist. that contains $\mathbb{P}$ with high prob.
- solve the *distributionally robust optimization (DRO)*

$$\widehat{J}_N := \inf_{x \in \mathbb{R}^n} \sup_{\mathbb{Q} \in \widehat{\mathcal{P}}_N} \mathbb{E}_{\mathbb{Q}}[f(x, \xi)]$$

Background: Ambiguity sets and DRO

Approach:

- find an *ambiguity set* $\widehat{\mathcal{P}}_N$ of prob. dist. that contains $\mathbb{P}$ with high prob.
- solve the *distributionally robust optimization (DRO)*

$$\widehat{J}_N := \inf_{x \in \mathbb{R}^n} \sup_{\mathbb{Q} \in \widehat{\mathcal{P}}_N} \mathbb{E}_{\mathbb{Q}}[f(x, \xi)]$$

Proposition (Adapted from Esfahani & Kuhn '17)

Let $\mathbb{P} \in \mathcal{M}(\mathbb{R}^m)$, *dist. with finite second moment. Let* $\widehat{\mathbb{P}}_N := \frac{1}{N} \sum_{k=1}^{N} \delta_{\widehat{\xi}^k}$,

$$\widehat{\mathcal{P}}_N := \mathcal{B}_{\epsilon_N(\beta)}(\widehat{\mathbb{P}}_N) = \{\mathbb{Q} \in \mathcal{M}(\mathbb{R}^m) \mid d_{W_2}(\widehat{\mathbb{P}}_N, \mathbb{Q}) \le \epsilon_N(\beta)\}.$$

Then, we have $\mathbb{P}^N(\mathbb{P} \in \widehat{\mathcal{P}}_N) \ge 1 - \beta$.

$$\epsilon_N(\beta) := \begin{cases} \left(\frac{\log(c_1 \beta^{-1})}{c_2 N}\right)^{1/\max\{4,m\}}, & \text{if } N \ge \frac{\log(c_1 \beta^{-1})}{c_2}, \\ \left(\frac{\log(c_1 \beta^{-1})}{c_2 N}\right)^{1/a}, & \text{if } N < \frac{\log(c_1 \beta^{-1})}{c_2}. \end{cases}$$

Background: Ambiguity sets and DRO

Approach:

- find an *ambiguity set* $\widehat{\mathcal{P}}_N$ of prob. dist. that contains $\mathbb{P}$ with high prob.
- solve the *distributionally robust optimization (DRO)*

$$\widehat{J}_N := \inf_{x \in \mathbb{R}^n} \sup_{\mathbb{Q} \in \widehat{\mathcal{P}}_N} \mathbb{E}_{\mathbb{Q}}[f(x, \xi)]$$

Theorem (Finite-sample guarantee)

Let $\mathbb{P} \in \mathcal{M}(\mathbb{R}^m)$ be a light-tailed distribution and $\beta \in (0, 1)$. Let $\widehat{\mathcal{P}}_N = \mathcal{B}_{\epsilon_N(\beta)}(\widehat{\mathbb{P}}_N)$. Then, the finite-sample guarantee holds:

$$\mathbb{P}^N \Big(\mathbb{E}_{\mathbb{P}}[f(\widehat{x}_N, \xi)] \leq \widehat{J}_N \Big) \geq 1 - \beta.$$

Background: Ambiguity sets and DRO

Approach:

- find an *ambiguity set* $\widehat{\mathcal{P}}_N$ of prob. dist. that contains $\mathbb{P}$ with high prob.
- solve the *distributionally robust optimization (DRO)*

$$\widehat{J}_N := \inf_{x \in \mathbb{R}^n} \sup_{\mathbb{Q} \in \widehat{\mathcal{P}}_N} \mathbb{E}_{\mathbb{Q}}[f(x,\xi)]$$

Theorem (Tractability (Adapted from Esfahani & Kuhn '17))

In addition to the previous hypotheses, assume f to be convex-concave. Then, solving DRO is same as

$$\inf_{\lambda \geq 0, x} \left\{ \lambda \epsilon_N^2(\beta) + \frac{1}{N} \sum_{k=1}^{N} \max_{\xi \in \mathbb{R}^m} \left(f(x,\xi) - \lambda \|\xi - \widehat{\xi}^k\|^2 \right) \right\}.$$

Distributed reformulation

Data-driven centralized problem

$$\inf_{\lambda \geq 0, x} \left\{ \lambda \epsilon_N^2(\beta) + \frac{1}{N} \sum_{k=1}^{N} \max_{\xi \in \mathbb{R}^m} \left(f(x, \xi) - \lambda \|\xi - \widehat{\xi}^k\|^2 \right) \right\} \qquad (\star)$$

Distributed problem: agent i's estimates x^i and λ^i

$$\min_{x_v, \lambda_v \geq \mathbf{0}_n} \quad \frac{\epsilon_N^2(\beta) \mathbf{1}_n^\top \lambda_v}{n} + \frac{1}{N} \sum_{k=1}^{N} \max_{\xi \in \mathbb{R}^m} \left(f(x^{v_k}, \xi) - \lambda^{v_k} \|\xi - \widehat{\xi}^k\|^2 \right) \qquad (\star\star)$$

$$\text{subject to} \quad \mathsf{L}\lambda_v = \mathbf{0}_n \quad \text{and} \quad (\mathsf{L} \otimes \mathsf{I}_d) x_v = \mathbf{0}_{nd}$$

(Here $x_v = (x^1; \dots; x^n)$, $\lambda_v = (\lambda^1; \dots; \lambda^n)$)

Lagrangian and its augmented version

$$L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) := \frac{\epsilon_N^2(\beta)\mathbf{1}_n^\top \lambda_{\mathsf{v}}}{n} + \sum_{k=1}^{N} \max_{\xi \in \mathbb{R}^m}\Big(f(x^{v_k}, \xi) - \lambda^{v_k}\|\xi - \widehat{\xi}^k\|^2\Big)$$
$$+ \nu^\top \mathsf{L}\lambda_{\mathsf{v}} + \eta^\top (\mathsf{L} \otimes \mathsf{I}_d) x_{\mathsf{v}}$$

Lagrangian and its augmented version

$$L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) := \frac{\epsilon_N^2(\beta)\mathbf{1}_n^\top \lambda_{\mathsf{v}}}{n} + \sum_{k=1}^{N} \max_{\xi \in \mathbb{R}^m} \Big(f(x^{v_k}, \xi) - \lambda^{v_k} \|\xi - \widehat{\xi}^k\|^2 \Big) + \nu^\top \mathsf{L}\lambda_{\mathsf{v}} + \eta^\top (\mathsf{L} \otimes \mathsf{I}_d) x_{\mathsf{v}}$$

Zero-duality gap:

$$\inf_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} \sup_{\nu, \eta} L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) = \sup_{\nu, \eta} \inf_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta).$$

Lagrangian and its augmented version

$$L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) := \frac{\epsilon_N^2(\beta)\mathbf{1}_n^\top \lambda_{\mathsf{v}}}{n} + \sum_{k=1}^{N} \max_{\xi \in \mathbb{R}^m}\Big(f(x^{v_k}, \xi) - \lambda^{v_k}\|\xi - \widehat{\xi}^k\|^2\Big) + \nu^\top \mathsf{L}\lambda_{\mathsf{v}} + \eta^\top(\mathsf{L} \otimes \mathsf{I}_d)x_{\mathsf{v}}$$

Zero-duality gap:

$$\inf_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} \sup_{\nu,\eta} L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) = \sup_{\nu,\eta} \inf_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta).$$

Augmented Lagrangian: (for better convergence properties)

$$L_{\mathrm{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) := L(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) + \frac{1}{2}x_{\mathsf{v}}^\top(\mathsf{L} \otimes \mathsf{I}_d)x_{\mathsf{v}} + \frac{1}{2}\lambda_{\mathsf{v}}^\top \mathsf{L}\lambda_{\mathsf{v}}$$

Lagrangian and its augmented version

$$L(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta) := \frac{\epsilon_N^2(\beta)\mathbf{1}_n^\top \lambda_{\mathrm{v}}}{n} + \sum_{k=1}^{N} \max_{\xi \in \mathbb{R}^m} \Big(f(x^{v_k}, \xi) - \lambda^{v_k} \|\xi - \widehat{\xi}^k\|^2 \Big) + \nu^\top \mathsf{L}\lambda_{\mathrm{v}} + \eta^\top (\mathsf{L} \otimes \mathsf{I}_d) x_{\mathrm{v}}$$

Zero-duality gap:

$$\inf_{x_{\mathrm{v}}, \lambda_{\mathrm{v}} \geq \mathbf{0}_n} \sup_{\nu, \eta} L(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta) = \sup_{\nu, \eta} \inf_{x_{\mathrm{v}}, \lambda_{\mathrm{v}} \geq \mathbf{0}_n} L(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta).$$

Augmented Lagrangian: (for better convergence properties)

$$L_{\mathrm{aug}}(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta) := L(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta) + \frac{1}{2} x_{\mathrm{v}}^\top (\mathsf{L} \otimes \mathsf{I}_d) x_{\mathrm{v}} + \frac{1}{2} \lambda_{\mathrm{v}}^\top \mathsf{L} \lambda_{\mathrm{v}}$$

Lemma (Lagrangians have same saddle points)

$(x_{\mathrm{v}}^*, \lambda_{\mathrm{v}}^*, \nu^*, \eta^*)$ saddle point of L over $(\mathbb{R}^{nd} \times \mathbb{R}^n_{\geq 0}) \times (\mathbb{R}^{n+nd})$ if and only if saddle point of L_{aug} over same domain

Modified Lagrangian

Get rid of the inner maximization

$$L_{\mathrm{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) = \max_{\{\xi^k\}} \tilde{L}_{\mathrm{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta, \{\xi^k\})$$

where

$$\tilde{L}_{\mathrm{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta, \{\xi^k\}) := \frac{\epsilon_N^2(\beta)\mathbf{1}_n^\top \lambda_{\mathsf{v}}}{n} + \sum_{k=1}^{N} \Big(f(x^{v_k}, \xi^k) - \lambda^{v_k} \|\xi^k - \widehat{\xi}^k\|^2 \Big)$$
$$+ \nu^\top \mathsf{L}\lambda_{\mathsf{v}} + \eta^\top (\mathsf{L} \otimes \mathsf{I}_d) x_{\mathsf{v}} + \frac{1}{2} x_{\mathsf{v}}^\top (\mathsf{L} \otimes \mathsf{I}_d) x_{\mathsf{v}} + \frac{1}{2} \lambda_{\mathsf{v}}^\top \mathsf{L} \lambda_{\mathsf{v}}$$

Modified Lagrangian

Get rid of the inner maximization

$$L_{\text{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) = \max_{\{\xi^k\}} \tilde{L}_{\text{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta, \{\xi^k\})$$

Saddle points of L_{aug} exists implying

$$\min_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} \max_{\nu, \eta} L_{\text{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta) = \max_{\nu, \eta} \min_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} L_{\text{aug}}(x_{\mathsf{v}}, \lambda_{\mathsf{v}}, \nu, \eta)$$

Substituting

$$\min_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} \max_{\nu, \eta} \max_{\{\xi^k\}} \tilde{L}_{\text{aug}}(\cdot) = \max_{\nu, \eta} \min_{x_{\mathsf{v}}, \lambda_{\mathsf{v}} \geq \mathbf{0}_n} \max_{\{\xi^k\}} \tilde{L}_{\text{aug}}(\cdot)$$

Modified Lagrangian

Get rid of the inner maximization

$$L_{\text{aug}}(x_{\text{v}}, \lambda_{\text{v}}, \nu, \eta) = \max_{\{\xi^k\}} \tilde{L}_{\text{aug}}(x_{\text{v}}, \lambda_{\text{v}}, \nu, \eta, \{\xi^k\})$$

Saddle points of L_{aug} exists implying

$$\min_{x_{\text{v}},\lambda_{\text{v}}\geq \mathbf{0}_n} \max_{\nu,\eta} L_{\text{aug}}(x_{\text{v}},\lambda_{\text{v}}, \nu, \eta) = \max_{\nu,\eta} \min_{x_{\text{v}},\lambda_{\text{v}}\geq \mathbf{0}_n} L_{\text{aug}}(x_{\text{v}}, \lambda_{\text{v}}, \nu, \eta)$$

Interchange and now,

$$\min_{x_{\text{v}},\lambda_{\text{v}}\geq \mathbf{0}_n} \max_{\nu,\eta,\{\xi^k\}} \tilde{L}_{\text{aug}}(\cdot) = \max_{\nu,\eta,\{\xi^k\}} \min_{x_{\text{v}},\lambda_{\text{v}}\geq \mathbf{0}_n} \tilde{L}_{\text{aug}}(\cdot)$$

Proposition (Correspondence between optima and saddle points)

If $((x_{\text{v}}^*, \lambda_{\text{v}}^*, \nu^*, \eta^*, \{(\xi^*)^k\})$ is saddle point of $\tilde{L}_{\text{aug}}$ over $\lambda_{\text{v}} \geq \mathbf{0}_n$, then $(x_{\text{v}}^*, \lambda_{\text{v}}^*, \nu^*, \eta^*)$ is primal-dual opt of $(\star\star)$

Distributed algorithm

Saddle-point dynamics for $\tilde{L}_{\text{aug}}$ is distributed

$$\begin{aligned}
\frac{dx_{\mathrm{v}}}{dt} &= -\nabla_{x_{\mathrm{v}}} \tilde{L}_{\text{aug}}(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta, \{\xi^k\}) \\
\frac{d\lambda_{\mathrm{v}}}{dt} &= [-\nabla_{\lambda_{\mathrm{v}}} \tilde{L}_{\text{aug}}(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta, \{\xi^k\})]^+_{\lambda_{\mathrm{v}}} \\
\frac{d\nu}{dt} &= \nabla_{\nu} \tilde{L}_{\text{aug}}(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta, \{\xi^k\}) \\
\frac{d\eta}{dt} &= \nabla_{\eta} \tilde{L}_{\text{aug}}(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta, \{\xi^k\}) \\
\frac{d\xi^k}{dt} &= \nabla_{\xi^k} \tilde{L}_{\text{aug}}(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta, \{\xi^k\}), \; \forall k \in \{1, \dots, N\}
\end{aligned}$$

Theorem (Asymptotic convergence)

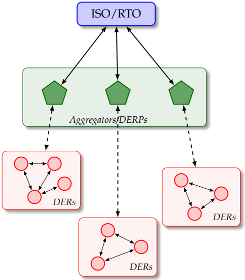
Assume $\exists$ *primal-dual opt.* $(x_{\mathrm{v}}^*, \lambda_{\mathrm{v}}^*, \nu^*, \eta^*)$ *with* $\lambda_{\mathrm{v}}^* \neq 0$*. Then, starting from* $\lambda_{\mathrm{v}}(0) \geq \mathbf{0}_n$*, trajectory converges asymptotically to saddle point of* $\tilde{L}_{\text{aug}}$ *over* $\lambda_{\mathrm{v}} \geq \mathbf{0}_n$ *and* $(x_{\mathrm{v}}, \lambda_{\mathrm{v}}, \nu, \eta)$ *converges to primal-dual optimizer*

[A. Cherukuri & J. Cortés, TAC, Submitted 2018]

Summary

In this talk:

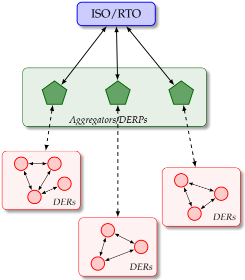
- hierarchical dispatch framework
- coordination of DERs in dispatch
- data-driven distributed optimization



Summary

In this talk:

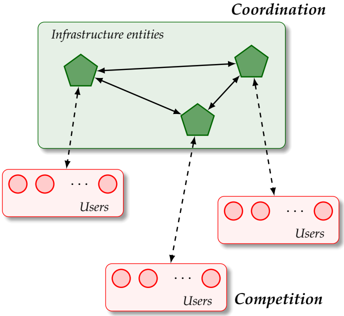
- hierarchical dispatch framework
- coordination of DERs in dispatch
- data-driven distributed optimization

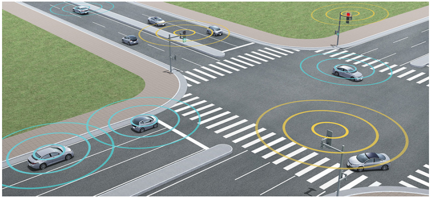


Future work:

- energy-efficient implementations of distributed algorithms
- data-driven chance-constrained optimization
  - finite and streaming data guarantees
  - distributed implementation

Intelligent transportation system





V2V and V2I communication

- For human driven vehicles
  - infrastructure entities coordinate and users compete
- For autonomous vehicles
  - infrastructure entities as well as users (vehicles) coordinate

Research directions:

- design of incentives using data: information or pricing
- data-driven coordination of traffic lights, ramp meters, variable speed limits

Thank you: Questions or Comments?