Optimization, Control, and Markets

6.S893: AI for Climate Action (Power & Energy Systems)

Spring 2026

Outline

Background: Electricity markets, optimal power flow, and ancillary services

Overview of ML for optimal power flow

Challenge highlight: Enforcing safety in ML

Outline

Background: Electricity markets, optimal power flow, and ancillary services

Overview of ML for optimal power flow

Challenge highlight: Enforcing safety in ML

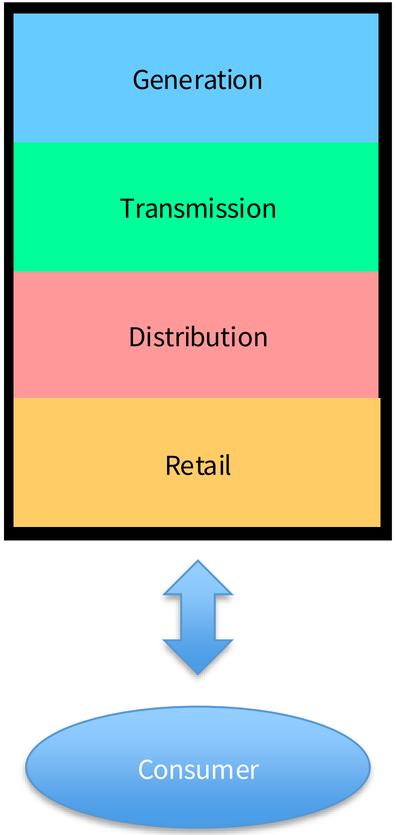
Traditional electric utility model

Monopoly – only supplier of electricity in a region ("service territory")

Vertically integrated – single organization performs all technical and business functions

Government-owed or heavily regulated

- Investor-owned utility (IOU) receives right to be monopoly supplier in return for letting regulator determine its rates ("rate of return regulation")



Adapted from Daniel Kirschen (2023)

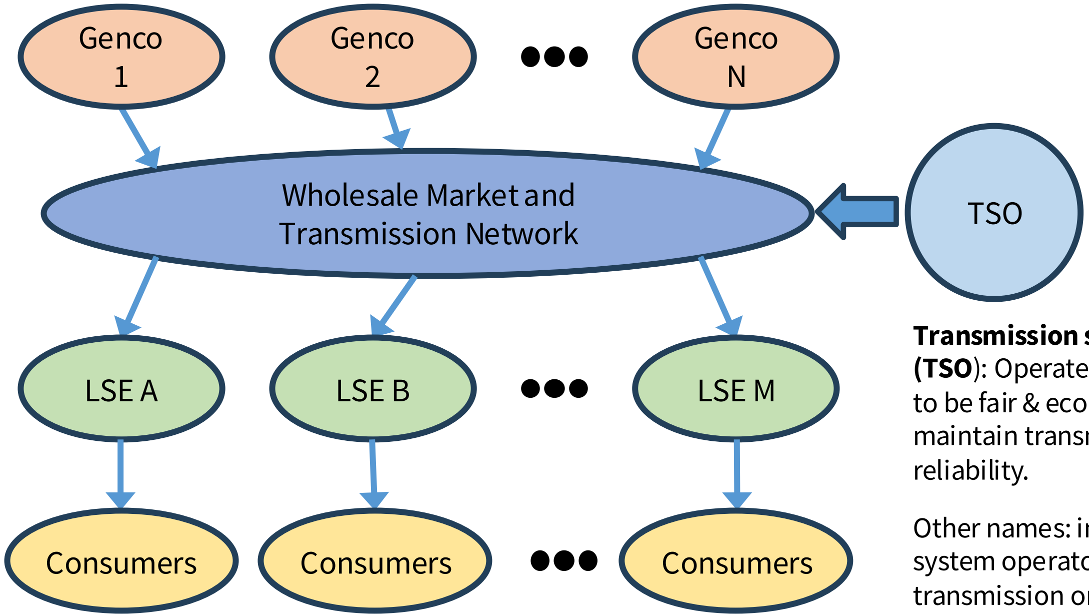
Deregulated model: Wholesale electricity market

Generating companies: Own and operate power plants, sell energy on wholesale market

Transmission companies: Own and operate transmission equipment

Load-serving entities: Buy energy on wholesale market, sell to consumers in monopoly service area

(Alternative model: Retail market)



Transmission system operator (TSO): Operate wholesale market to be fair & economically efficient, maintain transmission system reliability.

Other names: independent system operator (ISO), regional transmission organization (RTO)

Small consumers: Buy electricity from retailer. Distribution-connected.

Large consumers: May buy electrical energy directly through market, offer TSO ability to control their load, and face more complex rate structures. May be directly transmission-connected.

Figure from: Daniel Kirschen (2023)

Other important stakeholders

Regulators: Ensure fair markets/encourage competition, set rates for monopoly transmission and distribution companies, enforce reliability standards

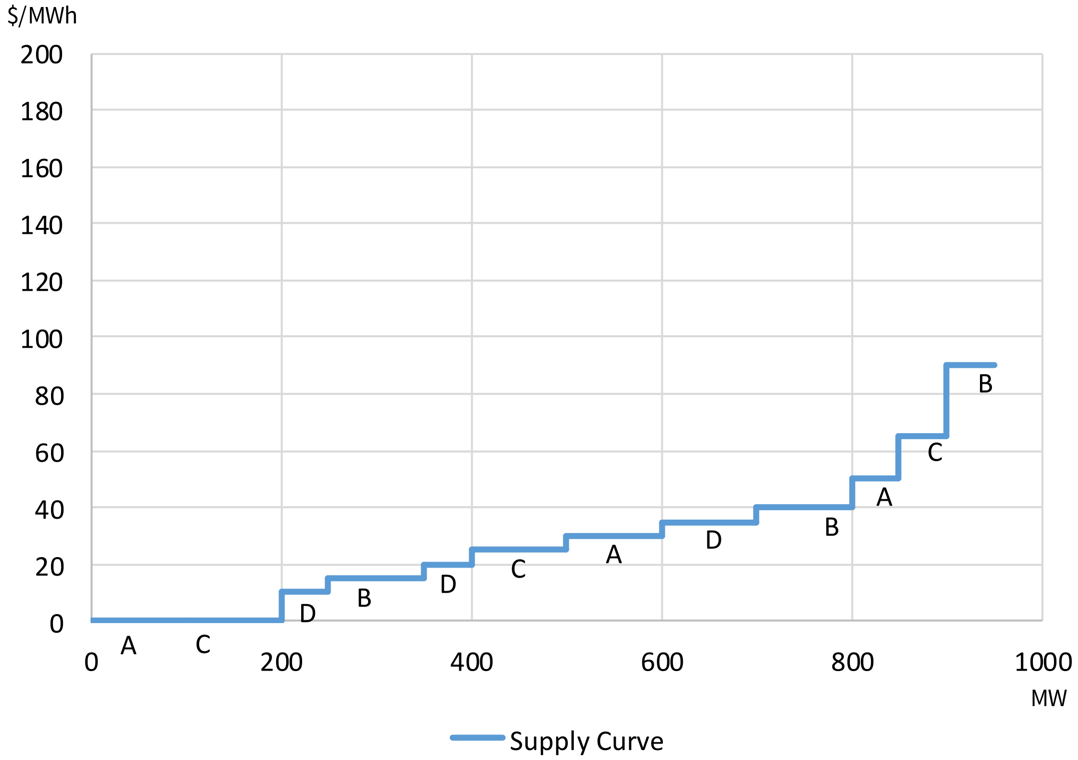
- **Federal Energy Regulatory Commission (FERC):** In US, regulates interstate transmission and wholesale sale of electricity and natural gas
- **North American Electric Reliability Corporation (NERC):** In North America, responsible for maintaining system standards and reliability
- **Public utility commissions:** In US, state-level entities regulating distribution networks and retail markets

Owners of storage systems: Can perform arbitrage in wholesale electricity market, provide ancillary services to TSO, or operate in conjunction with renewables

Electricity traders: May buy/sell in markets without owning any assets

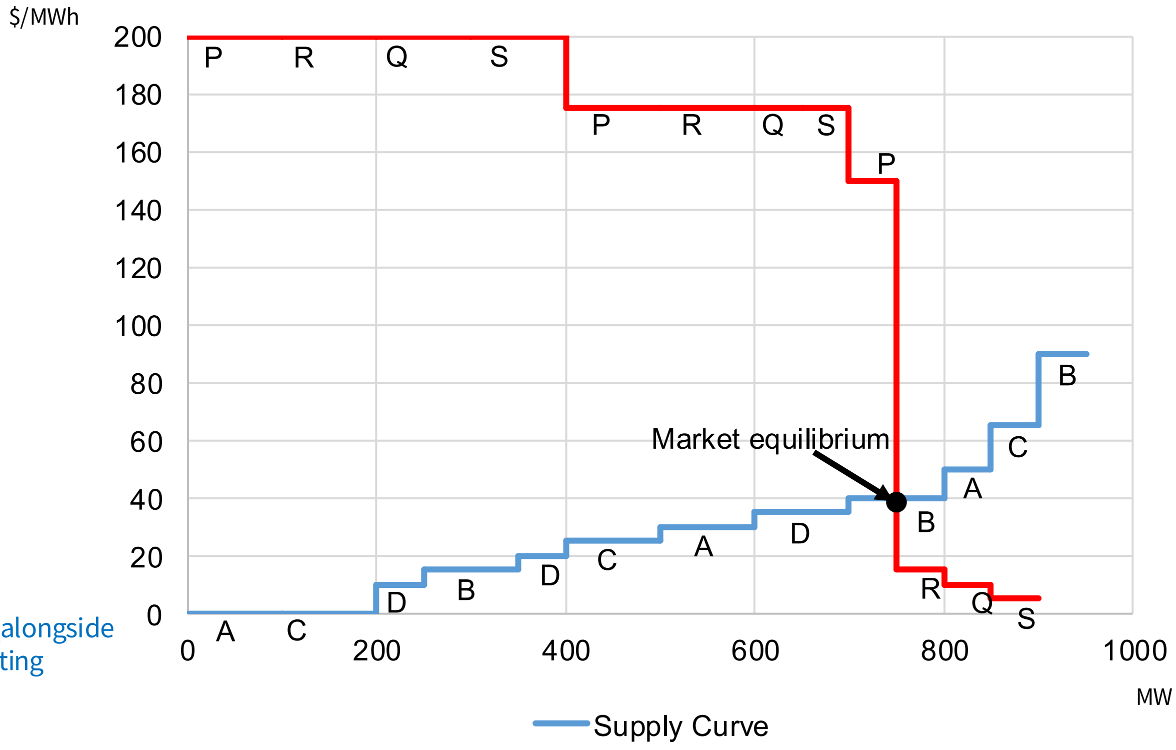
Merit order curve

Offers to Sell		
Genco	Quantity (MW)	Price ($/MWh)
A	100	0.00
	100	30.00
	50	50.00
B	100	15.00
	100	40.00
	50	90.00
C	100	0.00
	100	25.00
	50	65.00
D	50	10.00
	50	20.00
	100	35.00



Centralized energy market clearing

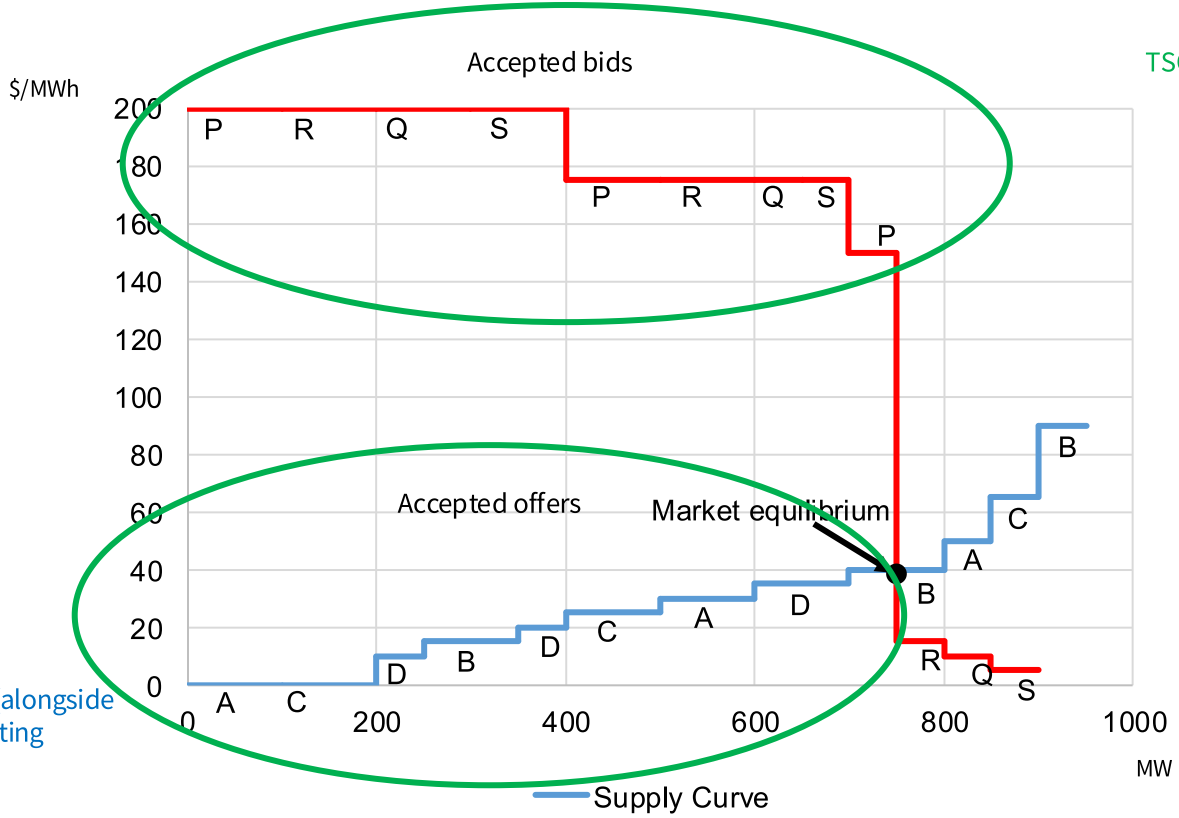
LSEs submit bids



Gencos submit offers (alongside info on physical operating parameters)

Centralized energy market clearing

LSEs submit bids



Gencos submit offers (alongside info on physical operating parameters)

TSO clears market

Market price: 40 $/MWh

Quantity traded: 750 MW

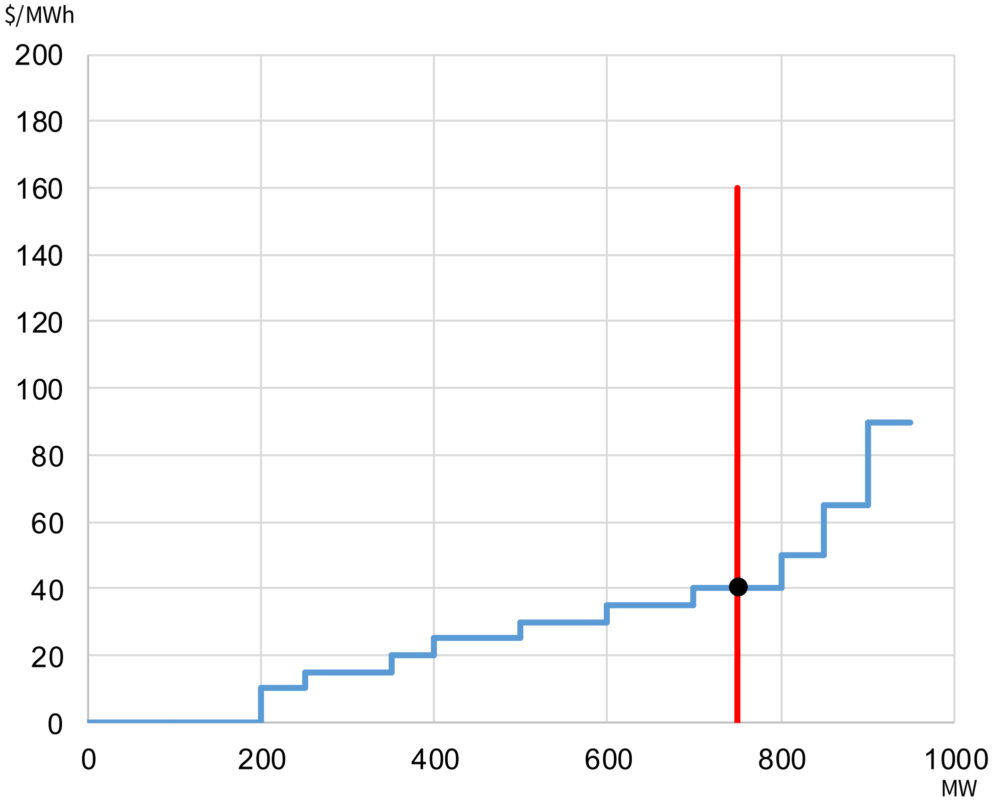
All energy traded is bought/sold at the market price (reflects marginal cost of generation)

Note: May need to make adjustments, e.g., if a generator doesn't hit $P^{\min}$

Replacing the demand curve by a load forecast

Some markets neglect the effect of the price on the demand

The demand curve is then replaced by a vertical line at the forecast value of the load



Economic dispatch

Goal: At a given point in time, determine how much power each generating unit should produce to supply the load at minimum cost

$$\underset{P_i \ \forall i \in G}{\text{minimize}} \sum_{i \in G} C_i(P_i)$$

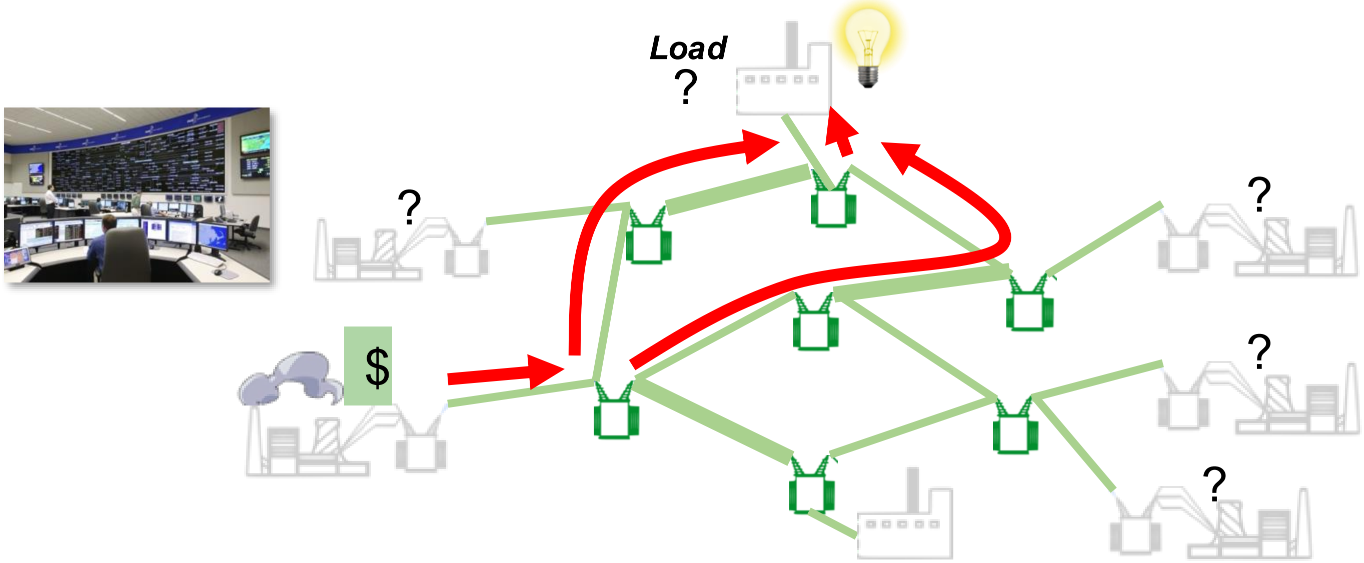
$$\text{subject to} \quad \sum_{i \in G} P_i = L$$

$$P_i^{\min} \leq P_i \leq P_i^{\max} \quad \forall i \in G$$

G: Set of generating units
C_i: Cost function for unit i
P_i: Power produced by unit i
L: Total load
$P_i^{\min}$: Min. gen. for unit i
$P_i^{\max}$: Max. gen. for unit i

For variable sources with zero fuel cost (e.g., wind and solar), often considered as "must-take" resources (e.g., modeled as negative load)

Recall: Grid as a network of generators and loads



Generators and loads are at **buses** (**nodes**)

Buses are connected by **power lines** (edges)

Effect of transmission capacity limits

Challenge: Pattern of generation and demand may result in violations of operating limits of the transmission network

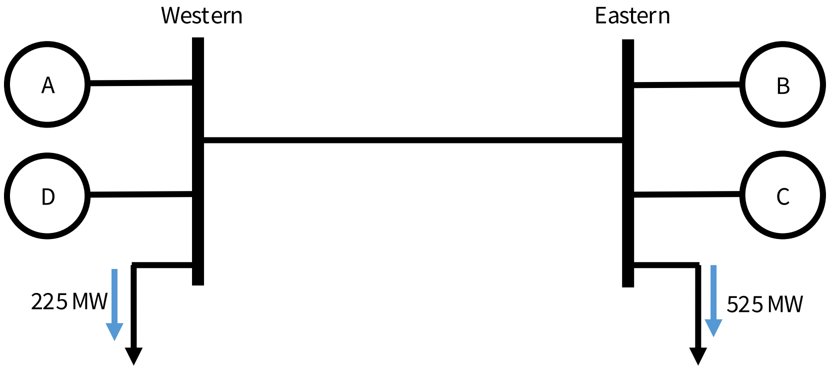
- Must modify results of market clearing to remove these violations
- Reject some cheaper offers in the "wrong" place, replace them with more expensive offers in the "right" place

Locational marginal price (LMP): Market price depends on the location

Example: Locational marginal pricing

Generators and loads separated in two regions connected by a single transmission line

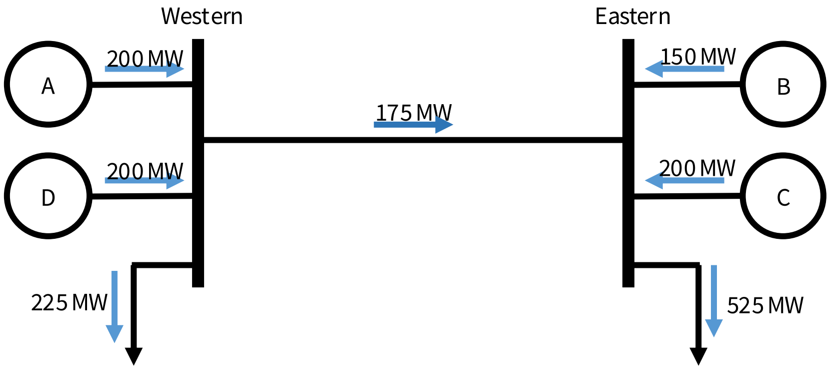
Assume, for now, a lossless system



Example: Locational marginal pricing, ctd.

Unconstrained case (no transmission limits or generator limits)

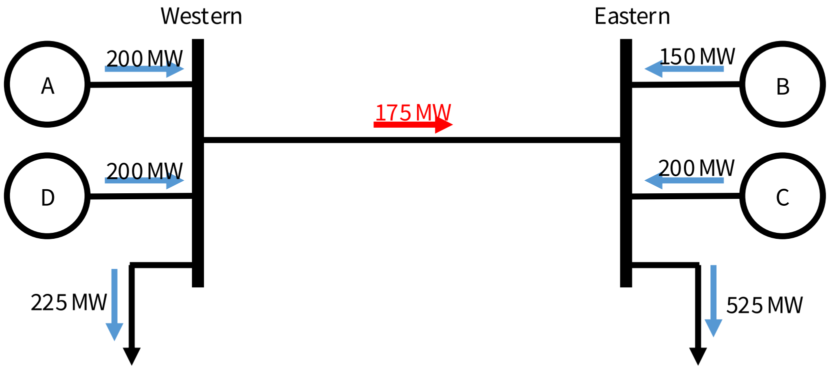
- Any generator can service any load
  - Total supply = total load
- Same market price everywhere
  - Equal to marginal cost of generation



Example: Locational marginal pricing, ctd.

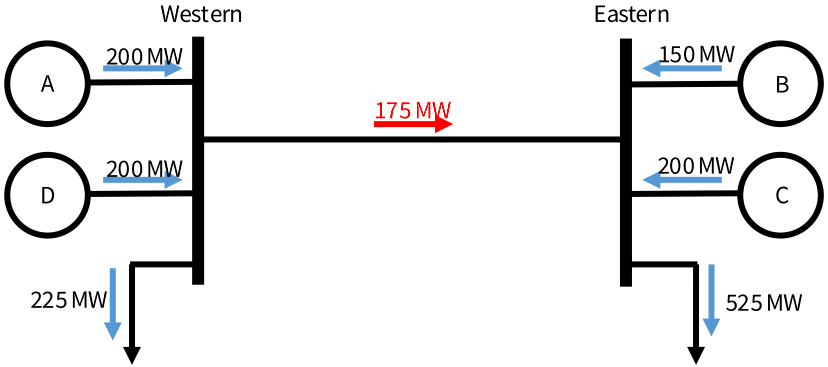
Constrained case: 50 MW rating for transmission line

- Violated by the unconstrained market clearing



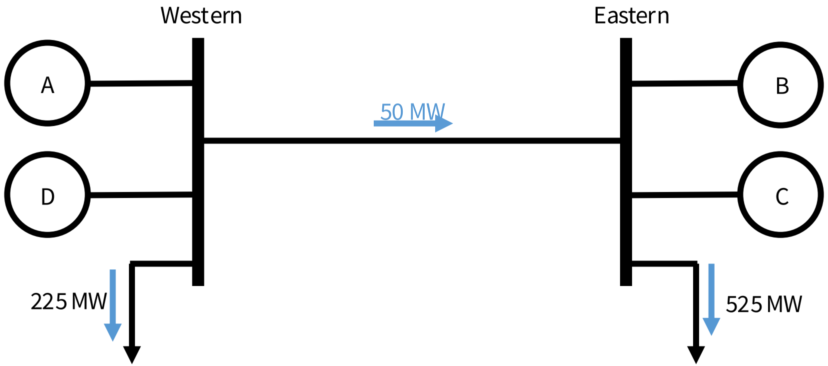
Example: Locational marginal pricing, ctd.

- To remove the overload:
  - Reduce the generation in the Western region by 125 MW
  - Increase the generation in the Eastern region by 125 MW

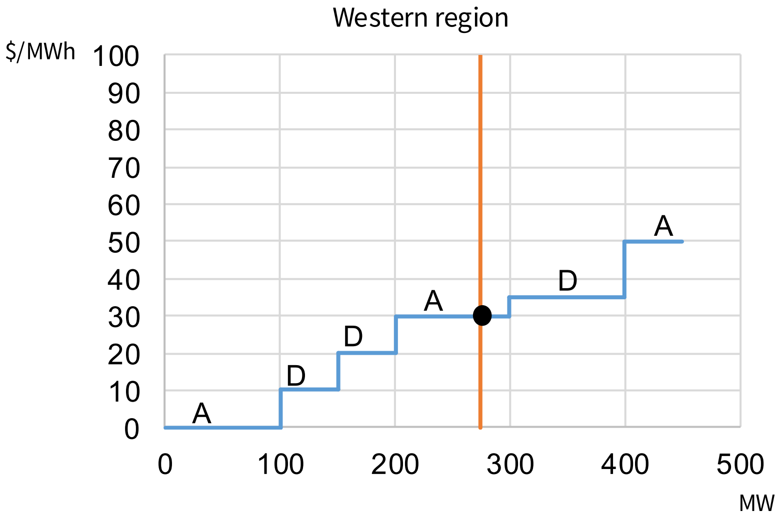


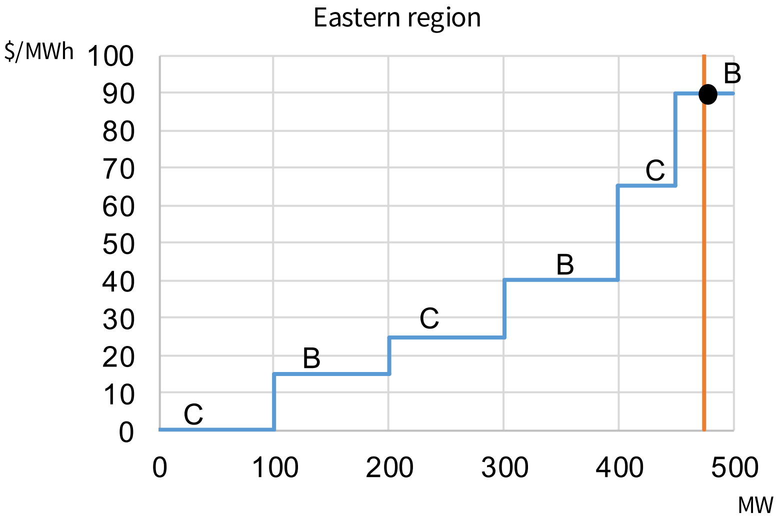
- A & D compete to supply 275 MW

- B & C compete to supply 475 MW



Example: Locational marginal pricing, ctd.





Market price in the Western region: 30 $/MWh

Market price in the Eastern region: 90 $/MWh

Locational marginal pricing

Generators are paid the LMP at the bus where they are connected

LSEs pay the LMP at the bus where they are connected

Prices will vary depending on which constraints are binding

Example with two zones and one transmission line

→ Easy to calculate the locational marginal prices

Larger, meshed transmission networks

→ Solve a linearized *optimal power flow*

→ Prices are the Lagrange multipliers of the nodal power balance constraints

Attributes of energy markets in practice

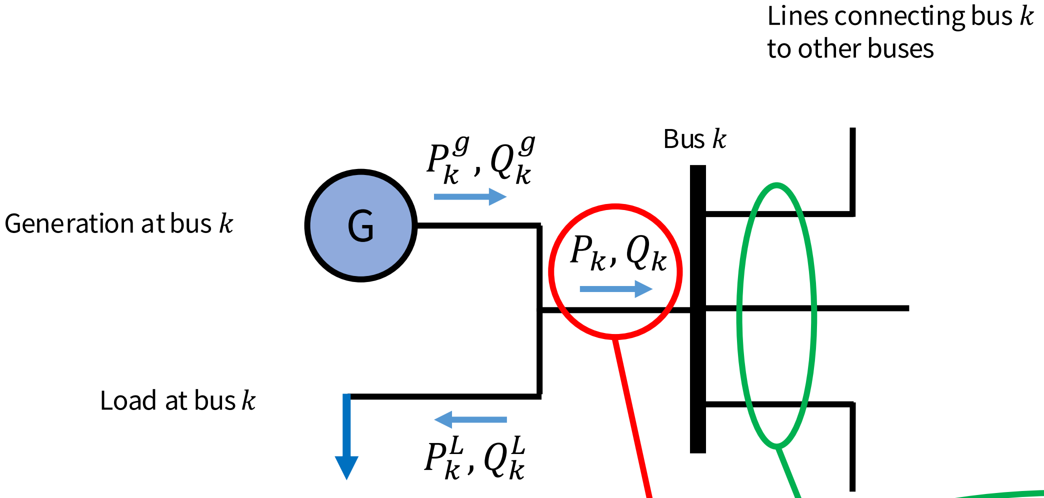
Common to have **two-settlement system**

- Day-ahead market (DAM)
  - Allow planning horizon (e.g., for units to turn on/off)
  - Based on *unit commitment*
- Real-time market (RTM)
  - Deal with real-time load forecast deviations, outages, etc.
  - Based on *economic dispatch*

Energy bought/sold via mix of bilateral contracts and wholesale markets

Many different financial products (futures/forward contracts, options)

Recall: Power flow equations



Active power injection at bus k: $P_k = P_k^g - P_k^L$

Reactive power injection at bus k: $Q_k = Q_k^g - Q_k^L$

$$P_k = \sum_{m=1}^{N} V_k V_m |Y_{km}| \cos(\varphi_k - \varphi_m - \delta_{km})$$

$$Q_k = \sum_{m=1}^{N} V_k V_m |Y_{km}| \sin(\varphi_k - \varphi_m - \delta_{km})$$

AC optimal power flow (AC-OPF)

$$\underset{\substack{P_i^g,\, Q_i^g\ \forall i \in G \\ V_k, \varphi_k\ \forall k \in N}}{\text{minimize}} \quad \sum_{i \in G} C_i(P_i^g)$$

subject to

$$P_k := P_k^g - P_k^L = \sum_{m=1}^{n} V_k V_m |Y_{km}| \cos(\varphi_k - \varphi_m - \delta_{km}) \quad \forall k \in N$$

Lagrange multipliers λ_k for $k \in B$ are nodal LMPs

$$Q_k := Q_k^g - Q_k^L = \sum_{m=1}^{n} V_k V_m |Y_{km}| \sin(\varphi_k - \varphi_m - \delta_{km}) \quad \forall k \in N$$

$$P_i^{\min} \leq P_i^g \leq P_i^{\max} \quad \forall i \in G$$

$$Q_i^{\min} \leq Q_i^g \leq Q_i^{\max} \quad \forall i \in G$$

$$V_k^{\min} \leq V_k \leq V_k^{\max} \quad \forall k \in N$$

$$\varphi_k^{\min} \leq \varphi_k \leq \varphi_k^{\max} \quad \forall k \in N$$

$$P_{km} := V_k V_m |Y_{km}| \cos(\varphi_k - \varphi_m - \delta_{km})$$

$$Q_{km} := V_k V_m |Y_{km}| \sin(\varphi_k - \varphi_m - \delta_{km})$$

$$\sqrt{P_{km}^2 + Q_{km}^2} \leq S_{km}^{\max} \quad \forall (k,m) \in E \cup E^R$$

AC optimal power flow (AC-OPF)

$$\underset{\substack{P_i^g,\, Q_i^g\ \forall i \in G \\ V_k, \varphi_k\ \forall k \in N}}{\text{minimize}} \quad \sum_{i \in G} C_i(P_i^g)$$

subject to

$$P_k := P_k^g - P_k^L = \sum_{m=1}^{n} V_k V_m |Y_{km}| \cos(\varphi_k - \varphi_m - \delta_{km}) \quad \forall k \in N$$

$$Q_k := Q_k^g - Q_k^L = \sum_{m=1}^{n} V_k V_m |Y_{km}| \sin(\varphi_k - \varphi_m - \delta_{km}) \quad \forall k \in N$$

$$P_i^{\min} \leq P_i^g \leq P_i^{\max} \quad \forall i \in G$$

$$Q_i^{\min} \leq Q_i^g \leq Q_i^{\max} \quad \forall i \in G$$

$$V_k^{\min} \leq V_k \leq V_k^{\max} \quad \forall k \in N$$

$$\varphi_k^{\min} \leq \varphi_k \leq \varphi_k^{\max} \quad \forall k \in N$$

$$P_{km} := V_k V_m |Y_{km}| \cos(\varphi_k - \varphi_m - \delta_{km})$$

$$Q_{km} := V_k V_m |Y_{km}| \sin(\varphi_k - \varphi_m - \delta_{km})$$

$$\sqrt{P_{km}^2 + Q_{km}^2} \leq S_{km}^{\max} \quad \forall (k,m) \in E \cup E^R$$

Lagrange multipliers λ_k for $k \in B$ are nodal LMPs

N : Set of all buses
$G \subset N$: Set of generator buses
C_i: Cost function for unit i

P_i^g: Active power produced by unit i
Q_i^g: Reactive power produced by unit i
V_k: Voltage magnitude at bus k
φ_k: Voltage angle at bus k

P_k, Q_k: Net active and reactive power injections at bus k
P_k^L, Q_k^L: Active and reactive power loads at bus k
N: Number of buses in the network
V_k: Magnitude of the voltage at node k
φ_k: Phase of the voltage at node k
$|Y_{km}|$: Magnitude of (k,m) term of admittance matrix Y
δ_{km}: Phase of (k,m) term of admittance matrix Y

$P_i^{\min}$, $P_i^{\max}$: Min./max active power gen. for unit i
$Q_i^{\min}$, $Q_i^{\max}$: Min./max reactive power gen. for unit i
$V_k^{\min}$, $V_k^{\max}$: Min./max voltage mag at bus k
$\varphi_k^{\min}$, $\varphi_k^{\max}$: Min./max voltage phase at bus k

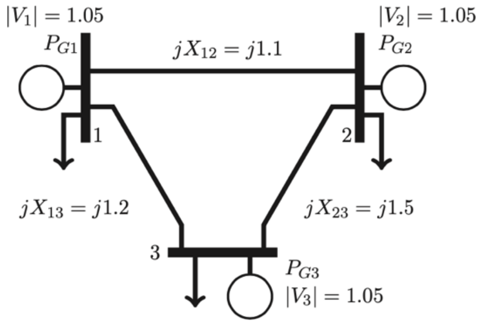
P_{km}: Active power flow on line (k,m)
Q_{km}: Reactive power flow on line (k,m)
$S_{km}^{\max}$: Flow limit on line (k,m)
E: Set of lines (E^R is its reverse)

AC-OPF is computationally expensive to solve

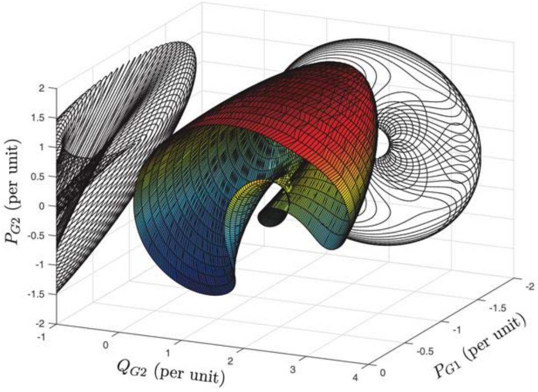
AC power flow (which appears in AC-OPF constraints) has a difficult structure

- <u>Nonlinearity:</u> AC power flow equations are nonlinear, which mean solutions may not be unique or may not exist
- <u>Nonconvexity:</u> AC power flow equations are nonconvex (and feasible space may be disconnected)

AC-OPF is therefore nonlinear, nonconvex, and proven to be NP-hard



(a) One-line diagram



(b) Feasible space

Figures: Molzahn, Daniel K., and Ian A. Hiskens. "A survey of relaxations and approximations of the power flow equations." *Foundations and Trends® in Electric Energy Systems* 4.1-2 (2019): 1-221.

AC-OPF is computationally expensive to solve (ctd.)

Case Name	Nodes	Edges	AC Time (sec.)
case1354_pegase	1354	1991	5
case1888_rte	1888	2531	9
case1951_rte	1951	2596	19
case2000_tamu	2000	3206	10
case2316_sdet	2316	3017	6
case2383wp_k	2383	2896	7
case2736sp_k	2736	3504	6
case2848_rte	2848	3776	17
case2853_sdet	2853	3921	9
case2868_rte	2868	3808	15
case2869_pegase	2869	4582	11
case3012wp_k	3012	3572	9
case3120sp_k	3120	3693	9
case3375wp_k	3374	4161	11
case4661_sdet	4661	5997	15
case6468_rte	6468	9000	64
case6470_rte	6470	9005	36
case6495_rte	6495	9019	69
case6515_rte	6515	9037	53
case9241_pegase	9241	16049	49
case10000_tamu	10000	12706	98
case13659_pegase	13659	20467	59

https://lanl-ansi.github.io/PowerModels.jl/stable/experiment-results/

Today, 50 years after the problem was formulated, we still do not have a fast, robust solution technique for the ACOPF. Finding a good solution technique for the ACOPF could potentially save tens of billions of dollars annually.

Source: Federal Energy Regulatory Commission (2020): https://ferc.gov/industries-data/electric/power-sales-and-markets/increasing-efficiency-through-improved-software-0

Variants of AC-OPF

Simplifications

- **DC-OPF** [commonly used]**:** Linear problem assuming $V = 1$,φ difference small, zero line resistance
- **Other relaxations**, e.g., semidefinite programming (SDP) relaxations

Harder/more realistic variants

- **N-k security-constrained OPF (SCOPF):** Robustness to contingencies (usually N-1)
- **Stochastic OPF:** Accounts for stochastic sources (e.g., loads, variable renewables)
- **Stability-constrained OPF:** Explicitly account for dynamic stability constraints

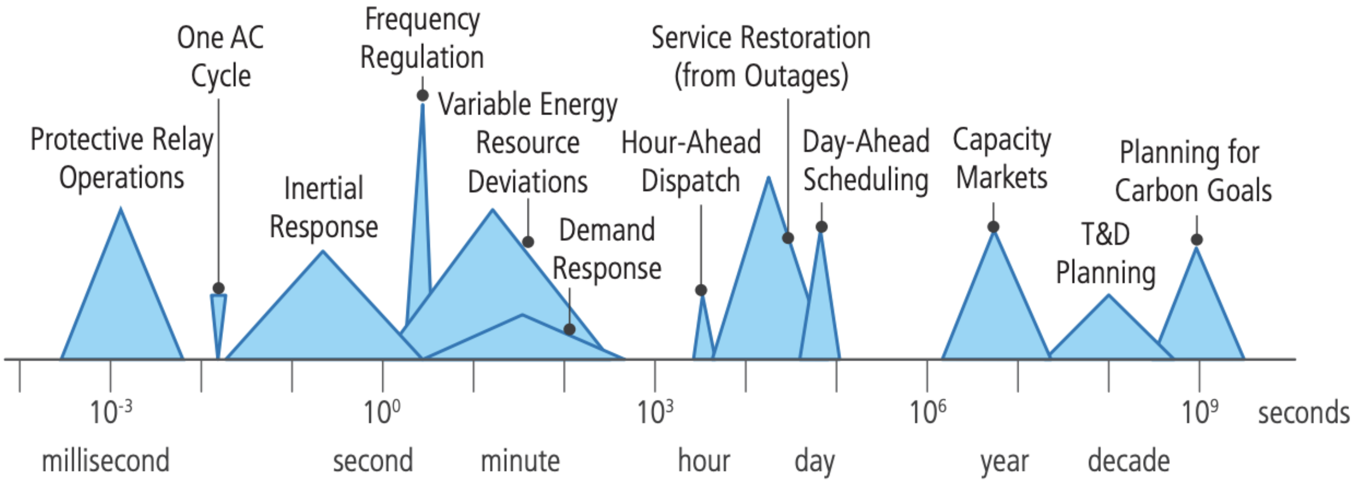
Problems built on top

- **Multi-period ACOPF:** Account for operation of generating units over time
- **Unit commitment:** Determine which generators to turn on/off, plus dispatch, over time
- **Capacity expansion planning:** Determine how to build/enhance the power system

Other related problems

- **Topology optimization:** Change the connectivity structure of the grid (bus splitting, line switching)

Figure S-5. System Reliability Depends on Managing Multiple Event Speeds



Capacity markets, day-ahead scheduling, and hour-ahead dispatch are well-understood tools for managing supply variability (mid-right axis). Beyond capacity contracts, traditional transmission and distribution (T&D) system long-term planning methods work to map and price investment requirements to ensure grid reliability (right end of axis). However, the widespread integration of variable energy resources significantly expands the time dimensions in which grid operators must function, ranging from hourly to minute to second intervals (mid-left axis). And, in a world of subsecond decision making (i.e., inertial response, one alternating current (AC) cycle, and protective relay operations), dispatch effectiveness will require the integration of automated grid management (left end of axis).

Figure source: US Department of Energy Quadrennial Energy Review (January 2017)

Ancillary services

TSO needs to ensure reliability of system

- Contingency reserve (spinning and non-spinning reserve)
  - Compensate providers for the opportunity cost of not providing energy
  - Opportunity for batteries
- Frequency control
  - Constant minor adjustments to power injections/extractions
  - Another opportunity for batteries
- Reactive power and voltage support
  - Providing reactive power causes losses and reduces opportunity to provide active power
- Black start capability
  - Need to be paid to be available

Other control of generators and loads

Maximizing generators' power production (done by generators)

- E.g., *maximum power point tracking (MPPT)* for solar panels – dynamically adjust I-V curve using power electronics
- E.g., *wind turbine control* and *wind farm control* – control the angle and orientation of wind turbines

Demand response and load flexibility (done by loads, often due to TSO incentives)

- Adjust amount and timing of loads for, e.g., peak load reduction, emissions reduction, or ancillary services

Outline

Background: Electricity markets, optimal power flow, and ancillary services

Overview of ML for optimal power flow

Challenge highlight: Enforcing safety in ML

Why use ML for AC-OPF?

Optimal power flow problems need to be solved at greater

- Speed – *due to variable renewables*
- Scale – *due to increasing number of devices*
- Fidelity (feasibility) – *due to, e.g., lowered overall system inertia*

Criteria not satisfied by, e.g., SDP relaxations or DC-OPF [B2021]

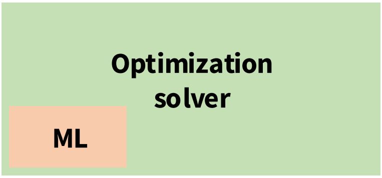
Principled use of ML can improve the solving of AC-OPF and other problems built upon it (e.g., SCOPF, stochastic OPF, unit commitment)

[B2021] Baker, Kyri. "Solutions of DC OPF are never AC feasible." *Proceedings of the Twelfth ACM International Conference on Future Energy Systems* (2021).

Modes for ML in AC-OPF

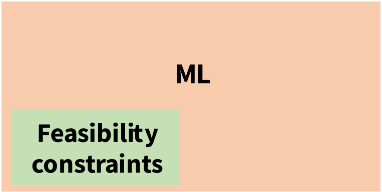
Speeding up existing AC-OPF solvers using ML:

- Learning warm start points
- Identifying inactive or redundant constraints



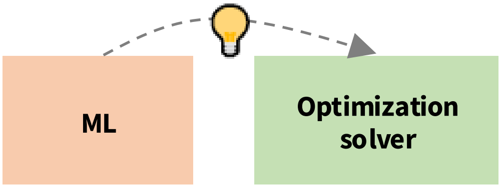
Learning feast, feasibility-constrained ML surrogates:

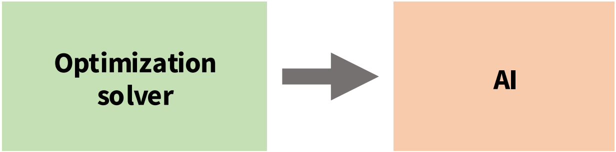
- Unconstrained ML model with post-hoc correction
- Feasibility enforcement internal to ML model



Other modes

- Fast solution techniques inspired by ML (but not using ML)
- Offline optimization + ML-based online actions





What makes an ML solver for AC-OPF “good” ?

Optimality: Is the objective value good?

Feasibility enforcement:

- Does the solver produce safe/feasible solutions when they exist?
- Does the solver behave appropriately when no feasible solution exists?

Speed and scalability:

- Is the solver fast enough for the use case it’s needed for?
- Can it solve realistic-scale problems?

Generalizability and robustness: Does the solver perform well under appropriate distributions of loads, generator bids, and topologies?

Other attributes, e.g.:

- Repeatable: Are the results reproducible, over and over again?
- Well-conditioned: Do similar inputs yield similar outputs?

Outline

Background: Electricity markets, optimal power flow, and ancillary services

Overview of ML for optimal power flow

Challenge highlight: Enforcing safety in ML

Safe ML for power systems

Soft-constrained methods: *Encourage* safety and feasibility

- E.g., Penalize constraint violations via the loss function

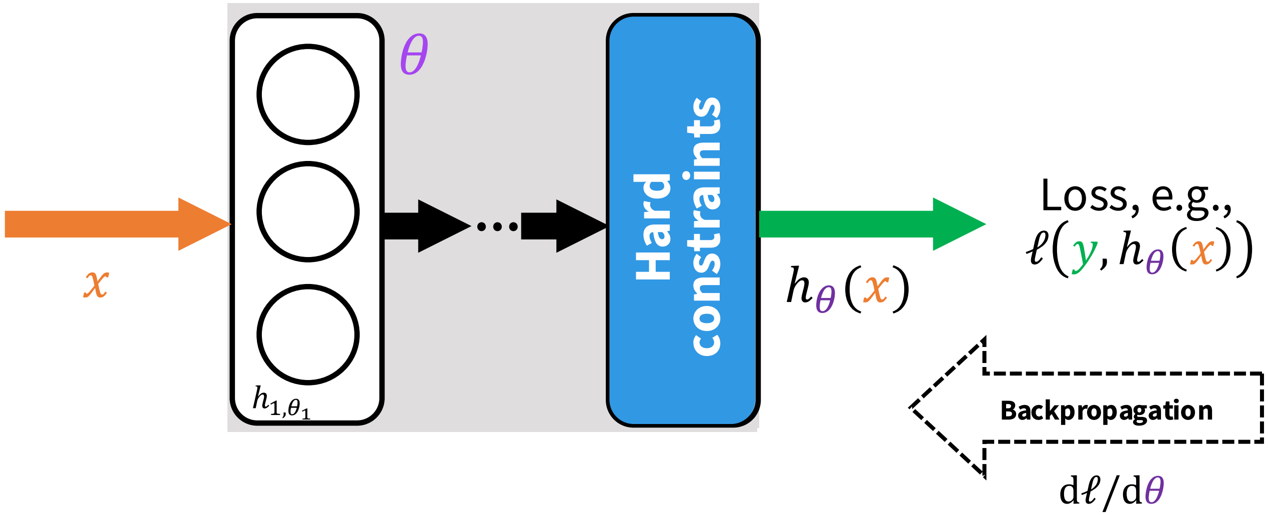
Hard constrained methods: *Enforce* safety and feasibility

- E.g., Design model architectures that enforce constraints by construction
- E.g., Enforce constraints during model optimization (projected gradient descent)
- E.g., Repair model feasibility post-hoc (e.g., via clipping or using a physical solver)

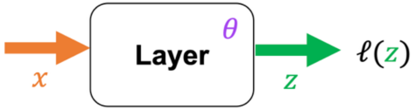
Verification and interpretability methods: *Understand* if/when a given model is safe

Example: Enforcing constraints within neural networks

Idea: Embed constraint enforcement as a differentiable "last layer" procedure



Explicit vs. implicit layers

	Explicit layer	**Implicit layer**
Forward pass: 	$z = f(x, \theta)$ [e.g., $z = \sigma(\theta^T x + \theta_0)$, unrolled optimization]	Find z such that $g(z, x, \theta) = 0$ [e.g., power flow, non-closed-form projections]
Backward pass: $\frac{\mathrm{d}\ell}{\mathrm{d}\theta} = \frac{\mathrm{d}\ell}{\mathrm{d}z^\star}\frac{\mathrm{d}z^\star}{\mathrm{d}\theta}$ $\quad \frac{\mathrm{d}\ell}{\mathrm{d}x} = \frac{\mathrm{d}\ell}{\mathrm{d}z^\star}\frac{\mathrm{d}z^\star}{\mathrm{d}x}$	$\frac{\mathrm{d}z^\star(x)}{\mathrm{d}x} = \frac{\mathrm{d}f(x,\theta)}{\mathrm{d}x}$	Find $\mathrm{d}z^\star(x)/\mathrm{d}x$ such that $\frac{\mathrm{d}g(z^\star(x), x, \theta)}{\mathrm{d}x} = 0$ by using implicit function theorem at a solution point

See also: Zico Kolter, David Duvenaud, and Matt Johnson. "Deep Implicit Layers - Neural ODEs, Deep Equilibirum Models, and Beyond." Tutorial at *NeurIPS 2020*. https://implicit-layers-tutorial.org/

Example: Implicit layer for QPs

Insight: Apply the implicit function theorem to the KKT optimality conditions

Example optimization problem (output: z)

$$\begin{aligned} \underset{z}{\text{minimize}} \quad & ½\, z^T Q z + q^T z \\ \text{subject to} \quad & Az = b \\ & Gz \le h \end{aligned}$$

Selected KKT optimality conditions

$$\begin{gathered} Qz^\star + q + A^T \nu^\star + G^T \lambda^\star = 0 \\ Az^\star - b = 0 \\ \text{diag}(\lambda^\star)(Gz^\star - h) = 0 \end{gathered}$$

Step 1: Apply implicit function theorem to the KKT conditions (generalized notation below)

$$\begin{bmatrix} Q & G^T & A^T \\ \text{diag}(\lambda^\star)G & \text{diag}(Gz^\star - h) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{d}z \\ \text{d}\lambda \\ \text{d}\nu \end{bmatrix} = - \begin{bmatrix} \text{d}Qz^\star + \text{d}q + \text{d}G^T\lambda^\star + \text{d}A^T\nu^\star \\ \text{diag}(\lambda^\star)\text{d}Gz^\star - \text{diag}(\lambda^\star)\text{d}h \\ \text{d}Az^\star - \text{d}b \end{bmatrix}$$

Generalized Jacobian of KKT conditions — Desired gradients — Gradients of problem parameters

Step 2: Use "Jacobian-vector trick" for efficient backpropagation

Brandon Amos and J. Zico Kolter. "OptNet: Differentiable optimization as a layer in neural networks." *ICML 2017.*
Priya L. Donti, Brandon Amos, and J. Zico Kolter. "Task-based end-to-end model learning in stochastic optimization." *NeurIPS 2017.*

Examples of other implicit layers

Insight: Apply implicit function theorem to equilibrium or optimality conditions (and use computational tricks to efficiently compute $\mathrm{d}\ell/\mathrm{d}\theta$ directly)

OptNet: Differentiable Optimization as a Layer in Neural Networks

Task-based End-to-end Model Learning in Stochastic Optimization

SATNet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver

Differentiable Submodular Maximization

End-to-End Differentiable Physics for Learning and Control

Filipe de A. Belbute-Peres
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
filiped@cs.cmu.edu

Kevin A. Smith
Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139
k2smith@mit.edu

Kelsey R. Allen
Brain and Cognitive Sciences
Massachusetts Institute of Technology

Joshua B. Tenenbaum
Brain and Cognitive Sciences
Massachusetts Institute of Technology

Differentiable Convex Optimization Layers

Neural Ordinary Differential Equations

Ricky T. Q. Chen*, Yulia Rubanova*, Jesse Bettencourt*, David Duvenaud
University of Toronto, Vector Institute
{rtqichen, rubanova, jessebett, duvenaud}@cs.toronto.edu

Takeaways

⚡ **Electricity markets, optimal power flow, and ancillary services**

- Energy market fundamentals: Economic dispatch, locational marginal pricing
- Formulation of AC optimal power flow (and its variants)
- Ancillary service markets and control help ensure system efficacy and reliability

🔎 **ML for AC-OPF:** Motivation, different approaches, and defining "success"

🏃 **Safe ML for power systems:**

- Soft-constrained vs. hard-constrained methods
- One approach: Implicit layers