Module 03
Introduction to System Optimization

Ahmad F. Taha

CE 4240/5240 — Introduction to Infrastructure Systems Engineering

Email: ahmad.taha@vanderbilt.edu

Webpage: https://ahmadstatespace.github.io/

VANDERBILT UNIVERSITY

Fall 2026

What you should be able to do

1. Turn an infrastructure decision into variables, an objective, and constraints.
2. Recognize convexity and identify LPs, QPs, QCQPs, and SOCPs.
3. Explain feasibility, optimality, and the value of extra capacity.
4. Interpret an algorithm's progress and verify its final answer.

Our recurring question: what does the mathematics mean for the system?

All numerical infrastructure cases in this module use illustrative data.

One module, five connected questions

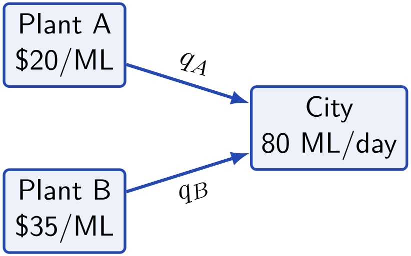
	Question	Infrastructure connection
1	What are we deciding?	Supply, flow, storage, and control
2	Is the model convex?	Mixing feasible operating plans
3	Which problem class fits?	LP, QP, QCQP, SOCP
4	How do we certify the answer?	KKT and capacity values
5	How do we compute it?	Short, practical algorithm toolkit

Formulate $\longrightarrow$ classify $\longrightarrow$ solve $\longrightarrow$ validate.

Start with a decision, not an equation

A water utility must supply 80 ML/day.

- q_A, q_B: treatment plant outputs.
- Cost: $20q_A + 35q_B$ dollars/day.
- Capacities: 60 and 50 ML/day.
- Demand must be met.



$$\min_{q_A, q_B} \; 20q_A + 35q_B \quad \text{s.t.} \quad q_A + q_B = 80, \quad 0 \le q_A \le 60, \quad 0 \le q_B \le 50.$$

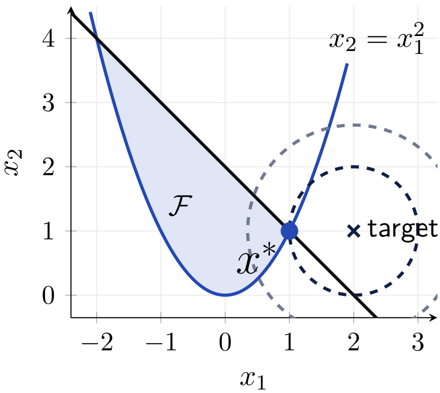
Use the cheaper plant first: $(q_A^*, q_B^*) = (60, 20)$.

A feasible region changes the best answer

$$\begin{aligned} \min_{x_1,x_2} \quad & (x_1-2)^2+(x_2-1)^2 \\ \text{s.t.} \quad & x_1^2 \le x_2, \\ & x_1+x_2 \le 2. \end{aligned}$$

- The unconstrained target $(2,1)$ is infeasible.
- The best feasible point is $(1,1)$, with objective value 1.

Contours show cost; the shaded region shows what is allowed.



A common language for every system

$$\begin{array}{llll} \min\limits_{x \in D} & f_0(x) & & \text{performance measure} \\ \text{s.t.} & g_i(x) \leq 0, & i = 1, \dots, m & \text{limits} \\ & h_j(x) = 0, & j = 1, \dots, p & \text{balances / physics} \end{array}$$

- $x \in \mathbb{R}^n$: decision vector; data are fixed for a particular solve.
- D: common domain of the objective and constraint functions.
- $\mathcal{F}$: points in D satisfying every constraint.
- $f^* = \inf_{x \in \mathcal{F}} f_0(x)$; an optimizer x^* attains that value.

Write units, signs, and time intervals before choosing a solver.

Feasible, infeasible, unbounded: different diagnoses

Status	Water-supply example	Interpretation
Feasible optimum	Demand 80; capacity $60 + 50$	At least one plan exists; minimize its cost.
Infeasible	Demand 120; capacity $60 + 50$	No plan meets every constraint.
Unbounded	Maximize sales revenue with no supply limit or cost	The model permits unlimited improvement.

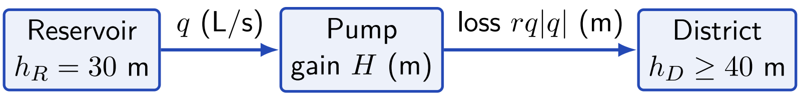
A solver failing to find a feasible point does not prove infeasibility.

Water systems: pressure, head, and flow

- **Head is energy per unit weight**, so its unit is meters:

$$h = z + \frac{p}{\rho g}, \qquad p = \rho g(h - z).$$

- z: elevation (m); p: gauge pressure (Pa). For water, $\rho \approx 1000$ kg/m^3 and $g \approx 9.81$ m/s^2: **10 m of pressure head** ≈ 98.1 **kPa**.
- Flow q is volume per time: 1 L/s $= 10^{-3}$ m^3/s. Hydraulic head h below omits velocity head $v^2/(2g)$.



$$\underbrace{q = d}_{\text{mass balance}}, \qquad \underbrace{h_D = h_R + H - rq|q|}_{\text{simplified steady energy balance}}.$$

With $r = 0.02$ m/(L/s)^2, $q = 20$ L/s and $H = 18$ m give $h_D = 40$ m.

Pump feasibility: formulate, solve, interpret

- **Data:** $h_R = 30$ m, minimum district head 40 m, pump limit 30 m; $r = 0.02$ m/(L/s)^2 and fixed $\eta = 0.8$.
- **Decision:** choose flow q, pump head H, and district head h_D. At fixed demand, minimizing H minimizes pump power.

Case A: demand 20 L/s

$$\begin{aligned} \min_{q,H,h_D} \quad & H \\ \text{s.t.} \quad & q = 20, \\ & h_D = 30 + H - 0.02q^2, \\ & h_D \geq 40, \quad 0 \leq H \leq 30. \end{aligned}$$

- $q^* = 20$, $H^* = 18$, $h_D^* = 40$.
- $P^* = \rho g (0.020)(18)/0.8$
  $= 4.41$ kW.

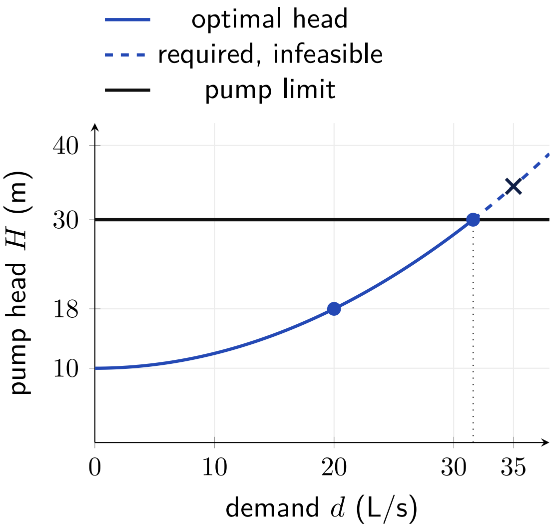
Case B: demand 35 L/s

$$\begin{aligned} \min_{q,H,h_D} \quad & H \\ \text{s.t.} \quad & q = 35, \\ & h_D = 30 + H - 0.02q^2, \\ & h_D \geq 40, \quad 0 \leq H \leq 30. \end{aligned}$$

- Would require $H \geq 34.5 > 30$ m.
- **Infeasible:** no optimizer or valid operating plan exists.

Substitute $q = d$ before solving: each case is an LP in H, h_D. An infeasible status is not a usable solution.

Demand versus capacity: where the pump runs out of head



- For $0 < d \leq 31.62$ L/s:

$$q^* = d, \quad H^* = 10 + 0.02d^2,$$
$$h_D^* = 40 \text{ m}.$$

- At $d = 35$, full pumping gives only $h_D = 35.5$ m.
- Capacity boundary:

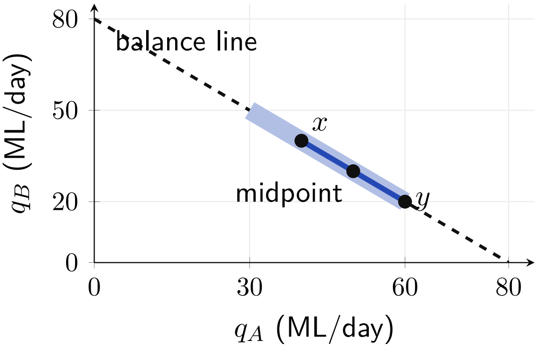
$$d_{\max} = \sqrt{\frac{30-10}{0.02}} \approx 31.62 \text{ L/s}.$$

Above the boundary, add capacity or revise service requirements; more solver iterations cannot help.

From a line to a segment to a convex set

- The **line** through distinct x, y is $z(\theta) = \theta x + (1-\theta)y$, $\theta \in \mathbb{R}$.
- Restrict $0 \le \theta \le 1$ to obtain the **line segment** joining them.
- A set C is **convex** if the entire segment stays in C for every $x, y \in C$.

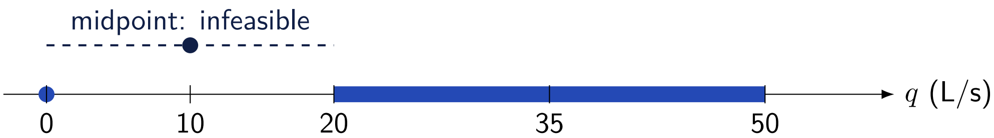
- Water plans: $q_A + q_B = 80$; $0 \le q_A \le 60$, $0 \le q_B \le 50$.
- $x = (40, 40)$, $y = (60, 20)$.
- $\theta = 1/2$ gives $(50, 30)$: demand and limits still hold.



Why on/off equipment breaks this picture

A pump is either off, or runs between 20 and 50 L/s:

$$q \in \{0\} \cup [20, 50].$$



A binary variable represents the choice:

$$20z \leq q \leq 50z, \qquad z \in \{0, 1\}.$$

Continuous equipment settings and discrete operating modes are different models.

Affine sets, halfspaces, and polyhedra have physical meanings

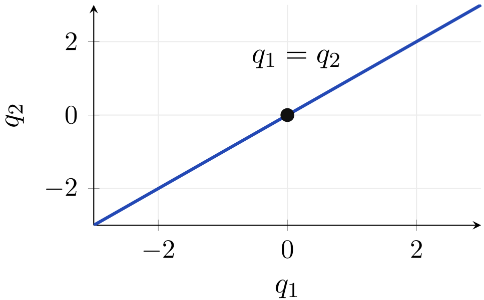
Geometry	Mathematical form	Infrastructure meaning
Subspace	$Bx = 0$	Balanced circulation without external supply
Affine set	$Bq = b$	Network flow with fixed injections
Halfspace	$a^T x \le b$	Budget, capacity, or emission limit
Polyhedron	$Ax \le b, \; Ex = d$	Intersection of finitely many linear limits and balances

Intersections of convex sets are convex: adding convex limits preserves convexity.

Subspaces and affine sets: specify the balance

Subspace: equal signed flows

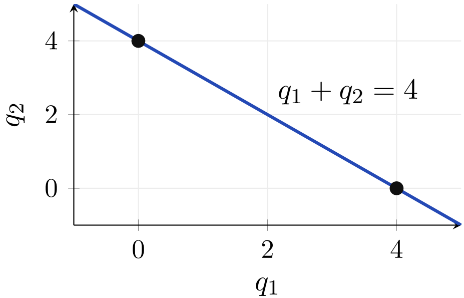
$$B = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad Bq = 0.$$



- Contains the origin; closed under sums and scalar multiplication.
- Negative values mean reverse flow.

Affine set: deliver 4 flow units

$$E = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad Eq = 4.$$

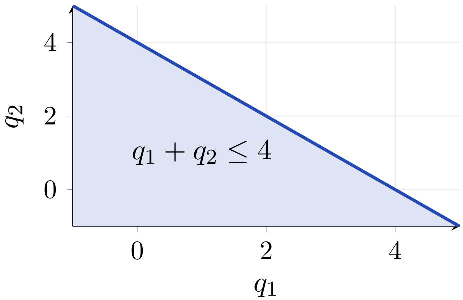


- A shifted subspace; need not contain zero.
- Adding nonnegativity leaves a segment.

Halfspaces intersect to make a polyhedron

One capacity halfspace

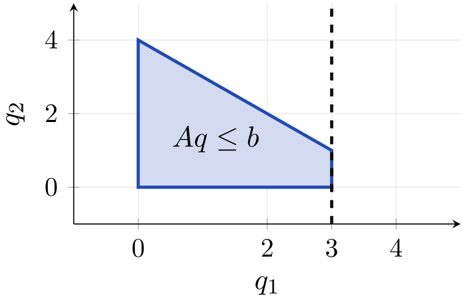
$$a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad a^T q \le 4.$$



- Shading continues beyond the plotting window.

Intersect with physical flow limits

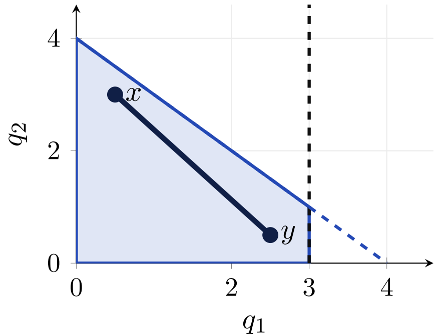
$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \\ 3 \\ 4 \end{bmatrix}.$$



Every face represents a specific balance or operating limit.

Why an intersection of convex sets stays convex

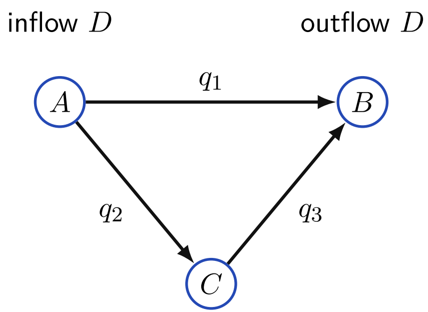
1. Choose any $x, y \in C_1 \cap C_2$.
2. Both points belong to each set. Convexity puts $\theta x + (1-\theta)y$ in each set for every $\theta \in [0,1]$.
3. Therefore the entire segment belongs to their intersection.



- Here $C_1 = \{q \geq 0 : q_1 + q_2 \leq 4\}$ and $C_2 = \{q : q_1 \leq 3\}$.
- Intersecting with $Eq = 4$ leaves the edge from $(0,4)$ to $(3,1)$: still convex.

The same argument works for any collection of convex constraints, even if the intersection is empty.

A transportation network becomes a matrix equation



Use $+1$ for an arc entering a node and -1 for an arc leaving:

$$\underbrace{\begin{bmatrix} -1 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}}_{B} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} -D \\ D \\ 0 \end{bmatrix}.$$

- Add $0 \leq q \leq \bar{q}$ for road capacities.
- One balance row is redundant: total inflow equals total outflow.

The same incidence matrix structure appears in water, gas, and power models.

Convex costs: the chord lies above the curve

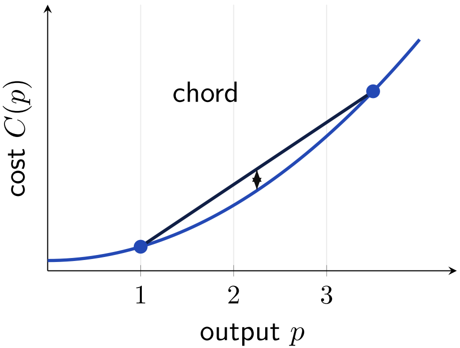
A function f on a convex domain is convex if

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y).$$

Example: generator cost

$$C(p) = ap^2 + bp + c, \qquad a \ge 0.$$

Strict inequality for distinct points and $0 < \theta < 1$ defines strict convexity.



Gradients and Hessians tell us about curvature

For differentiable f on an open convex domain:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \quad \Longleftrightarrow \quad f \text{ is convex.}$$

If f is twice continuously differentiable:

$$\nabla^2 f(x) \succeq 0 \text{ everywhere} \quad \Longleftrightarrow \quad f \text{ is convex.}$$

For dispatch, $C(p) = 0.025p_1^2 + 20p_1 + 0.05p_2^2 + 25p_2$:

$$\nabla C = \begin{bmatrix} 0.05p_1 + 20 \\ 0.10p_2 + 25 \end{bmatrix}, \qquad \nabla^2 C = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.10 \end{bmatrix} \succ 0.$$

Positive definite curvature gives strict convexity here.

Build larger convex models from simple pieces

Convexity-preserving operation	Engineering example
Nonnegative weighted sum	$\sum_i w_i(q_i - q_i^{\mathrm{ref}})^2$, $w_i \geq 0$
Affine substitution	$\|Hx - y\|_2^2$: fit a linear sensor model
Pointwise maximum	$\max_t p_t$: peak electrical demand
Convex sublevel set	$\{x : f(x) \leq c\}$: a convex resource budget

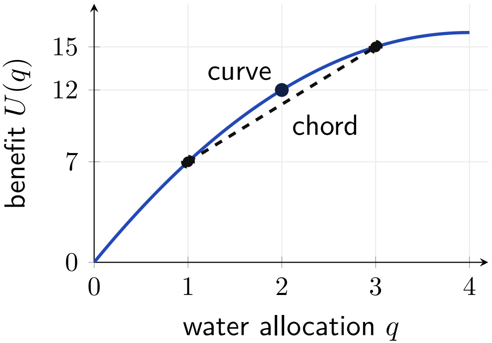
A product of decision variables is not automatically convex.

Concavity: diminishing returns make a benefit flatten

- A concave function reverses the convexity inequality:

$$U(\theta x+(1-\theta)y) \geq \theta U(x)+(1-\theta)U(y).$$

- Water-service benefit: $U(q) = 8q - q^2$, $0 \leq q \leq 4$, in illustrative benefit units.
- $U'(q) = 8 - 2q$ decreases: each extra unit of water adds less benefit.



- With unit supply cost $2q$, maximize $U(q) - 2q = 6q - q^2$.
- Equivalently minimize the convex function $q^2 - 6q$: $q^* = 3$.

Maximize a concave benefit or minimize a convex cost over a convex feasible set.

The convex model starter pack

$$\min_x f_0(x) \qquad \text{s.t.} \quad g_i(x) \leq 0, \quad Ax = b.$$

1. The objective f_0 is convex on its domain.
2. Each inequality function g_i is convex.
3. Equality constraints are affine in this standard representation.

$P^2 + Q^2 \leq S^2$, fixed S	Convex disk: equipment capability
$P^2 + Q^2 = S^2$	Circle: generally nonconvex
$q^2 \leq \Delta h / r$, fixed $r > 0$	Convex inequality
$q^2 = \Delta h / r$	Nonlinear equality: generally nonconvex

Local, global, and unique are different claims

- A **local** minimum beats nearby feasible points.
- A **global** minimum beats all feasible points.
- For a convex objective over a convex set, every local minimum is global.
- Strict convexity gives at most one optimizer; it does not guarantee existence.

Infrastructure contrast:

- Fixed mode convex dispatch: local optimality is enough.
- Pump switching or nonlinear hydraulic equalities: local methods may depend on initialization.

Convexity helps certification; model size, scaling, and sparsity still affect computation.

Four problem classes we will use

Class	Defining structure	Infrastructure example
LP	Linear objective; affine constraints	Supply allocation and min-cost flow
QP	Quadratic objective; affine constraints	Economic dispatch and tracking control
QCQP	Quadratic objective and inequalities	Dispatch with a quadratic budget
SOCP	Linear objective; affine and norm-cone constraints	Robust capacity and facility location

QP and QCQP labels alone do not imply convexity: inspect the quadratic matrices.

LP: allocate limited infrastructure resources

$$\min_x \quad c^T x$$
$$\text{s.t.} \quad Ax \le b, \quad Ex = d, \quad \ell \le x \le u.$$

For our two water plants:

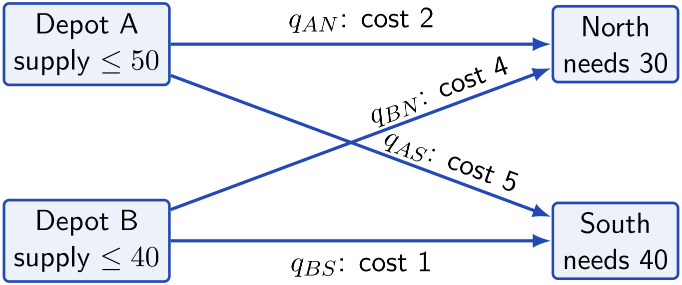
$$\min\ 20q_A + 35q_B \quad \text{s.t.} \quad q_A + q_B = 80, \quad 0 \le q_A \le 60, \quad 0 \le q_B \le 50.$$

Eliminate $q_B = 80 - q_A$:

$$30 \le q_A \le 60, \qquad C = 2800 - 15q_A.$$

$q_A^* = 60$, $q_B^* = 20$, and $C^* = 1900$ dollars/day.

A network LP: emergency water deliveries



Flow units: ML/day; costs: dollars/ML. No route-capacity limits in this example.

$$\min \; 2q_{AN} + 5q_{AS} + 4q_{BN} + q_{BS}.$$

What balances and capacity inequalities belong under this objective?

Emergency deliveries: formulate and solve

$$
\begin{aligned}
\min_{q \geq 0} \quad & 2q_{AN} + 5q_{AS} + 4q_{BN} + q_{BS} \\
\text{s.t.} \quad & q_{AN} + q_{AS} \leq 50, \\
& q_{BN} + q_{BS} \leq 40, \\
& q_{AN} + q_{BN} = 30, \\
& q_{AS} + q_{BS} = 40.
\end{aligned}
$$

Cheapest routes can serve every district:

$$
(q_{AN}, q_{AS}, q_{BN}, q_{BS}) = (30, 0, 0, 40), \qquad C^* = 100.
$$

Discussion: if B loses 10 units of supply, the cost rises to 140. Why?

Answer: move 10 units of South's delivery from B to A; extra cost $10(5-1) = 40$.

LP reformulations: slacks and absolute deviations

Slack = unused capacity

$$a^T x \leq b \quad \Longleftrightarrow \quad a^T x + s = b,\ s \geq 0.$$

- Plant capacity: $q \leq 60$ becomes $q + s = 60$.
- At $q = 45$, $s = 15$ ML/day is unused capacity.
- A free variable can be written $x = x^+ - x^-$ with $x^+, x^- \geq 0$.

Absolute error = two linear bounds

$$|q - r| \leq t \quad \Longleftrightarrow \quad -t \leq q - r \leq t.$$

- Track target supply $r = 50$ ML/day.
- At $q = 45$, the smallest feasible t is 5.
- Minimizing $\sum_i t_i$ models total absolute deviation exactly.

Auxiliary variables make useful engineering costs fit a linear model.

The epigraph: lift the cost into one extra variable

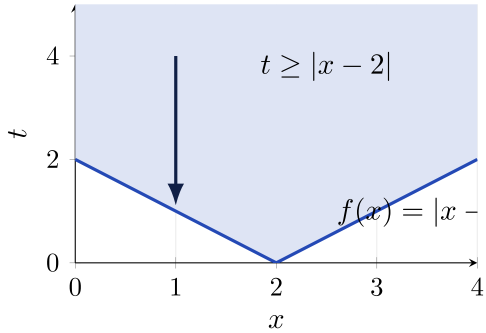
- The **epigraph** of f is the region on or above its graph:

$$\operatorname{epi} f = \{(x,t) : t \geq f(x)\}.$$

- Minimize its height:

$$\min_x f(x) \iff \min_{x,t}\{t : f(x) \leq t\}.$$

- If $t > f(x)$, lower t without changing x; at an attained optimum $t = f(x)$.



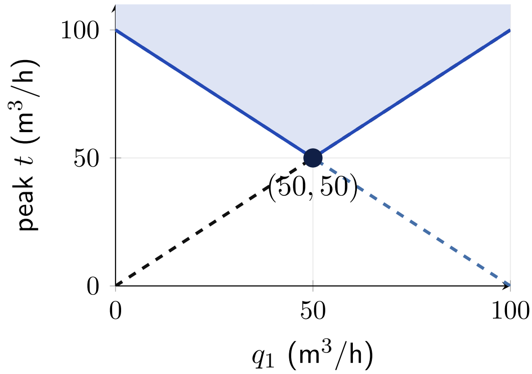
For $f(x) = \max_i(a_i^T x + b_i)$, impose every $a_i^T x + b_i \leq t$: the reformulation is an LP.

An epigraph LP balances two pumping periods

- A tank shifts supply between two one-hour periods; total pumped volume is $100\ \text{m}^3$.
- Choose $q_1, q_2 \geq 0$ (m^3/h) to minimize peak flow.

$$\begin{aligned} \min_{q_1,q_2,t} \quad & t \\ \text{s.t.} \quad & q_1 + q_2 = 100, \\ & q_1 \leq t, \quad q_2 \leq t, \\ & q_1, q_2 \geq 0. \end{aligned}$$

- Since $100 = q_1 + q_2 \leq 2t$, $t \geq 50$. Equal pumping attains the bound.



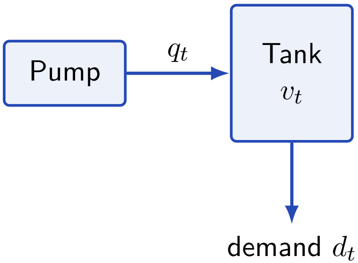
One variable t bounds both periods; minimizing it finds the lowest possible peak.

Assumes sufficient tank capacity and compatible initial/terminal storage; the next model includes those constraints.

Peak shaving with water storage is an LP

Over T one-hour intervals, choose pumping q_t and tank volumes v_t:

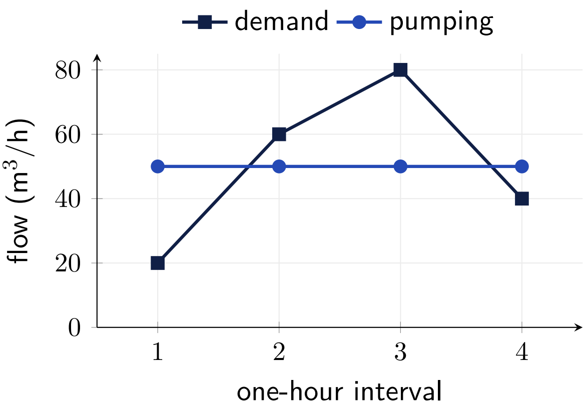
$$
\begin{aligned}
\min \quad & \sum_{t=1}^{T} \pi_t \kappa q_t \Delta t + \gamma P \\
\text{s.t.} \quad & v_{t+1} = v_t + \Delta t (q_t - d_t), \\
& \kappa q_t \le P, \quad 0 \le q_t \le \bar{q}, \\
& 0 \le v_t \le V, \quad v_{T+1} = v_1.
\end{aligned}
$$



- Fixed head and efficiency give power $p_t = \kappa q_t$.
- π_t: energy price.
- γ: peak-charge coefficient for the horizon.

v_1 is given. Bound all storage states, including the terminal state. Flow in m^3/h, volume in m^3, and κ in $\mathrm{kWh}/\mathrm{m}^3$.

Storage moves pumping away from the demand peak



Demand: $(20, 60, 80, 40)$.
Pump a constant 50 m^3/h; start with $v_1 = 20$ m^3.

$$(v_1, \ldots, v_5) = (20, 50, 40, 10, 20).$$

A 60 m^3 tank makes this feasible.

Peak pumping falls from 80 to 50: a 37.5% reduction.

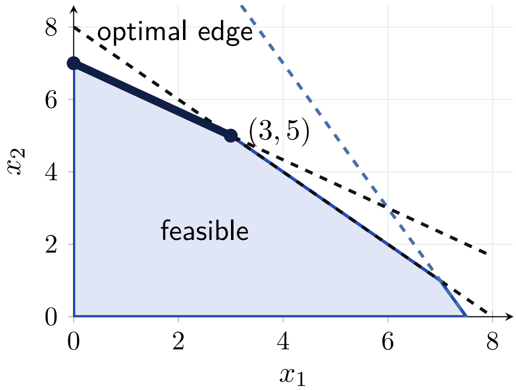
With equal energy prices and $\gamma > 0$, 50 is the minimum possible peak: it equals average demand.

An LP can have an entire edge of optimal solutions

$$\begin{aligned} \max_{x\geq 0} \quad & 4x_1 + 6x_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 8, \\ & 2x_1 + x_2 \leq 15, \\ & 2x_1 + 3x_2 \leq 21. \end{aligned}$$

- The objective equals $2(2x_1 + 3x_2)$.
- Every point from $(0,7)$ to $(3,5)$ is optimal, with value 42.

Non-strict convexity permits multiple optima; it does not require them.



QP: penalize deviations smoothly

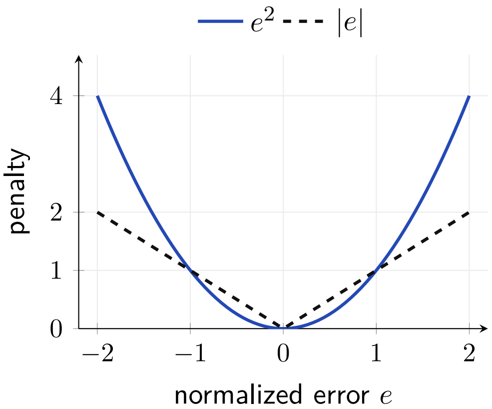
$$\begin{aligned} \min_{x} \quad & \tfrac{1}{2}x^T Q x + c^T x + r \\ \text{s.t.} \quad & Ax \le b, \quad Ex = d. \end{aligned}$$

Assume $Q = Q^T$. Then $\nabla f = Qx + c$ and $\nabla^2 f = Q$.

- $Q \succeq 0$: convex QP.
- $Q \succ 0$: strictly convex objective; at most one optimizer.
- $Q = 0$: LP.
- Indefinite Q: generally a nonconvex QP.

QP costs naturally express tracking, smoothing, and increasing marginal costs.

Why quadratic costs appear across infrastructure

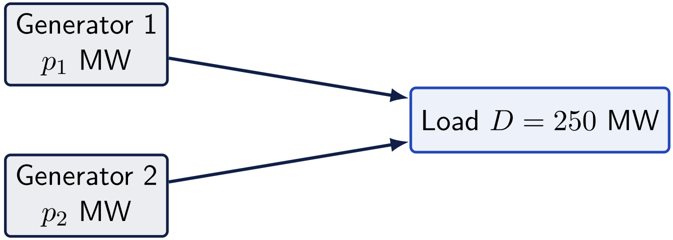


- **Tracking:** $(h - h_{\text{ref}})^2$ penalizes pressure deviations in either direction.
- **Smoothing:** $(u_t - u_{t-1})^2$ discourages abrupt pump or ramp-meter changes.
- **Marginal cost:** $ap^2 + bp$, $a > 0$, makes additional generation progressively more expensive.

- Doubling an error quadruples its squared penalty; gradients are smooth and linear in the error.
- A second-order Taylor model is locally quadratic. Use unit-aware weights; squared loss is sensitive to outliers.

Quadratics combine useful behavior with simple derivatives and efficient convex models when $Q \succeq 0$.

Economic dispatch: which generator should work harder?



$$\min_{p_1, p_2} \quad C_1(p_1) + C_2(p_2)$$

$$\text{s.t.} \quad p_1 + p_2 = 250, \quad 0 \le p_i \le 300,$$

$$C_1(p_1) = 0.025 p_1^2 + 20 p_1,$$

$$C_2(p_2) = 0.05 p_2^2 + 25 p_2.$$

Costs are dollars/hour. This aggregate model ignores transmission losses and line constraints.

Dispatch solution: equalize marginal costs

Eliminate $p_2 = 250 - p_1$:

$$C(p_1) = 0.075p_1^2 - 30p_1 + 9375, \qquad 0 \le p_1 \le 250.$$

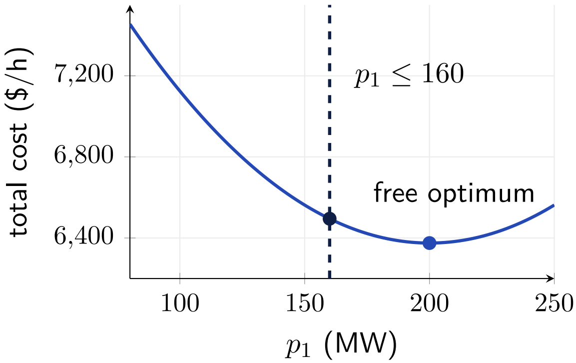
$$C'(p_1) = 0.15p_1 - 30 = 0 \quad \Rightarrow \quad (p_1^*, p_2^*) = (200, 50).$$

Plan	p_1 (MW)	p_2 (MW)	Cost ($/h)
Equal sharing	125	125	6796.875
Optimal dispatch	200	50	6375.000

$$C_1'(200) = 30 = C_2'(50) \quad \text{dollars/MWh.}$$

Equal marginal costs, not equal outputs: savings of 421.875 dollars/hour.

A capacity limit changes the dispatch



Now cap Generator 1 at 160 MW.

$$(p_1^*, p_2^*) = (160, 90),$$
$$C^* = 6495.$$

Marginal costs are unequal:

$$C_1'(160) = 28, \qquad C_2'(90) = 34.$$

One more MW of cheap capacity is worth about 6 dollars/hour locally.

A sensor-estimation QP for a water network

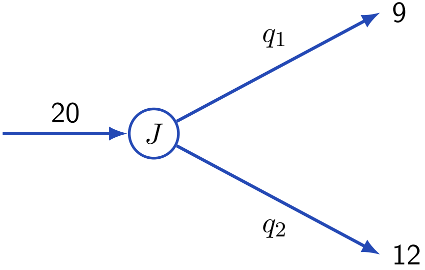
Two branch flows q_1, q_2 have readings $9, 12$; the inlet meter reads 20.

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 9 \\ 12 \\ 20 \end{bmatrix}.$$

$$\min_{q \geq 0} \; \|Hq - y\|_2^2 .$$

Equal sensor weights give

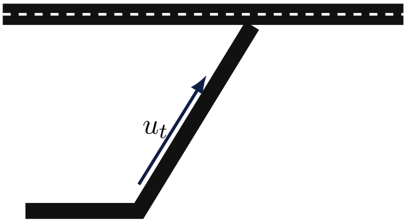
$$q^* = \left(\frac{26}{3}, \frac{35}{3}\right).$$



All readings use the same flow unit. The meters disagree: $9 + 12 \neq 20$.

A QP reconciles noisy measurements; weights express sensor confidence.

Transportation control: ramp metering as a QP



Use a local linear prediction model:

$$x_{t+1} = Ax_t + Bu_t + Ed_t, \quad x_t = (n_t, z_t).$$

Choose admitted ramp flow u_t:

$$\min \sum_{t=0}^{T-1} \left[\left\| x_t - x^{\mathrm{ref}} \right\|_Q^2 + r(u_t - u_{t-1})^2 \right].$$

Add linear queue, density, and actuator limits; $Q \succeq 0$, $r > 0$.

Measure, optimize, apply the first control, and repeat: model predictive control.

$\|e\|_Q^2 = e^T Q e$; x_0, u_{-1} are given. The linear model is a local approximation to traffic dynamics.

QCQP: put quadratic limits into the model

$$
\begin{aligned}
\min_{x} \quad & \tfrac{1}{2}x^T Q_0 x + c_0^T x + r_0 \\
\text{s.t.} \quad & \tfrac{1}{2}x^T Q_i x + c_i^T x + r_i \le 0, \quad i = 1, \ldots, m, \\
& Ax = b.
\end{aligned}
$$

A sufficient standard-form convexity test is

$$
Q_i \succeq 0 \qquad i = 0, 1, \ldots, m.
$$

- Quadratic objectives and constraints: a QCQP.
- A QP has only affine constraints.
- Quadratic equalities and indefinite matrices require separate analysis.

The direction of the inequality matters as much as the square.

QCQP example: dispatch with an emission budget

Use normalized outputs $p_1 + p_2 = 10$, with $p_1, p_2 \geq 0$. Generator 1 is cheaper but more emission-intensive:

$$\begin{aligned} \min \quad & p_1 + 3p_2 \\ \text{s.t.} \quad & p_1 + p_2 = 10, \\ & 0.1p_1^2 + 0.025p_2^2 \leq 4. \end{aligned}$$

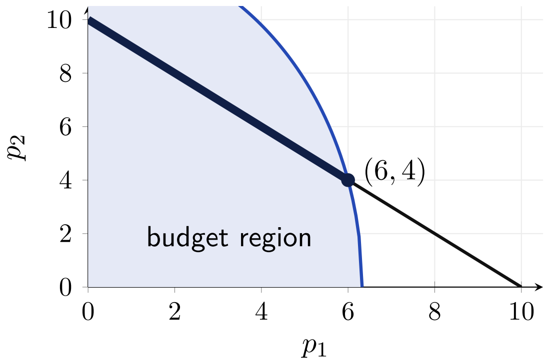
Substitute $p_2 = 10 - p_1$:

$$0.125p_1^2 - 0.5p_1 - 1.5 \leq 0 \quad \Rightarrow \quad 0 \leq p_1 \leq 6.$$

$(p_1^*, p_2^*) = (6, 4)$, cost 18, and the emission budget binds.

Illustrative convex emission proxy in normalized units; physical emission curves must be fitted to data.

Visualizing the emission–cost tradeoff



Navy: dispatch plans satisfying both demand and the budget.
Along the demand line,

$$C = 30 - 2p_1.$$

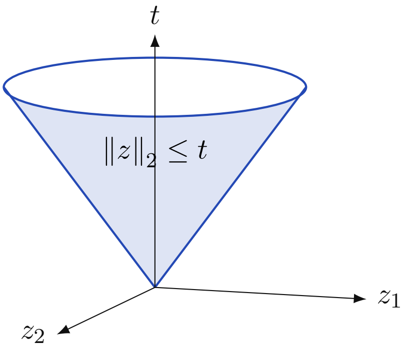
More output from Generator 1 reduces cost, until the curved limit stops us.

Discussion: where does the optimum move if the budget increases?

SOCP: Euclidean norms become constraints

The second-order cone is $\mathcal{Q} = \{(t, z) : \|z\|_2 \leq t\}$.

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \\ & Fx = g. \end{aligned}$$



The right side must be nonnegative; the norm inequality enforces this.
Affine inequalities may also be included.

Conic modeling reference: MOSEK Modeling Cookbook, Chapter 3.

Power example: inverter capability is a disk

An inverter's active and reactive power satisfy

$$P^2 + Q^2 \le S_{\max}^2$$

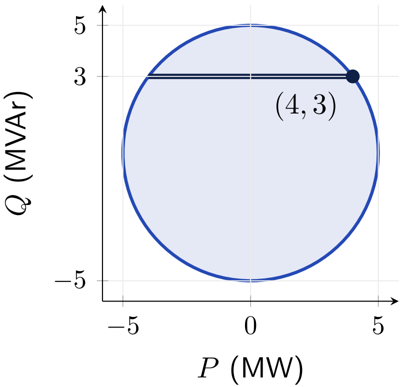
$$\iff$$

$$\|(P, Q)\|_2 \le S_{\max}.$$

For $S_{\max} = 5$ MVA and required $Q = 3$ MVAr:

$$|P| \le \sqrt{25 - 9} = 4 \text{ MW}.$$

Reactive-power support uses capacity that could deliver active power.



This capability constraint can be written as a convex quadratic inequality or as an SOC constraint.

Facility location: place an emergency service hub

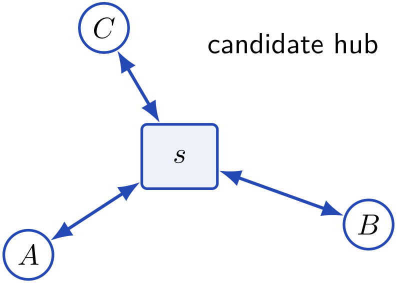
Let $a_i \in \mathbb{R}^2$ be district locations and $w_i > 0$ their service weights.

$$\min_{s \in \mathbb{R}^2} \sum_i w_i \|s - a_i\|_2 .$$

Introduce a distance variable t_i:

$$\min_{s,t} \sum_i w_i t_i \quad \text{s.t.} \quad \|s - a_i\|_2 \le t_i.$$

Convex land-use constraints on s can be added.



Euclidean distance gives an SOCP; road travel time may require a different model.

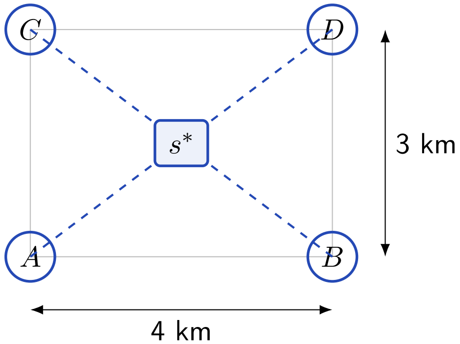
A worked hub location example

Four equally weighted districts:

$$(0,0),\ (4,0),\ (0,3),\ (4,3) \quad \text{km.}$$

Their center is $s^* = (2, 1.5)$. Each distance is 2.5 km, so total distance is 10 km. Why optimal? Pair opposite corners. The triangle inequality gives at least 5 km per pair.

The center attains the lower bound $5 + 5 = 10$.



Robust capacity: plan for uncertain coefficients

A feeder's approximate loading is $a^T x$, where x is a vector of controllable loads. Suppose

$$a = \bar{a} + Eu, \qquad \|u\|_2 \leq 1.$$

Require the capacity limit for every allowed uncertainty:

$$(\bar{a} + Eu)^T x \leq b \qquad \forall \|u\|_2 \leq 1.$$

The largest uncertain contribution is

$$\max_{\|u\|_2 \leq 1} u^T E^T x = \left\| E^T x \right\|_2 .$$

The robust constraint is $\bar{a}^T x + \left\| E^T x \right\|_2 \leq b$: an SOCP constraint.

An ellipsoidal uncertainty set gives this exact reformulation. See CVXPY's SOCP example.

A nominally feasible charging plan can fail the robust test

Two controllable charging loads $x_1, x_2 \geq 0$ share a normalized feeder capacity of 10:

$$\bar{a} = (1, 1), \qquad E = \mathrm{diag}(0.2, 0.1).$$

The robust condition is

$$x_1 + x_2 + \sqrt{(0.2x_1)^2 + (0.1x_2)^2} \leq 10.$$

Plan	Nominal load	Worst-case load	Robustly feasible?
$(5, 5)$	10.000	11.118	No
$(4.4, 4.4)$	8.800	9.784	Yes

The uncertainty margin is decision-dependent, not a fixed percentage added afterward.

Convex QCQPs can be represented with cones

For a quadratic inequality with $Q = R^T R \succeq 0$,

$$x^T Q x + a^T x + b \leq 0 \quad \Longleftrightarrow \quad \|Rx\|_2^2 \leq t, \quad t = -a^T x - b.$$

The squared-norm bound has an ordinary SOC representation:

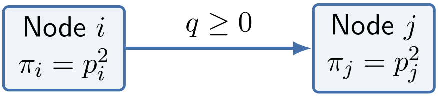
$$\boxed{\left\| \begin{bmatrix} 2Rx \\ t-1 \end{bmatrix} \right\|_2 \leq t+1.}$$

Squaring gives $4\,\|Rx\|_2^2 + (t-1)^2 \leq (t+1)^2$, hence $\|Rx\|_2^2 \leq t$.

Epigraph the quadratic objective too: LP $\subseteq$ convex QP $\subseteq$ convex QCQP $\subseteq$ SOCP-representable.

Use consistently scaled variables. The SOC also enforces $t \geq 0$. These are representation inclusions, not solver speed rankings.

Gas networks: a cone can relax nonlinear physics



Simplified steady passive pipe, fixed flow direction, no compressor or linepack.

A normalized quadratic pressure-drop model is

$$q^2 = k(\pi_i - \pi_j), \quad k > 0.$$

Relax the equality:

$$q^2 \leq k(\pi_i - \pi_j) = t.$$

This is conic-representable:

$$\|(2q, t-1)\|_2 \leq t + 1.$$

The inequality enlarges the feasible set; it is not automatically exact.

Teaching model: all variables in the cone are normalized. Pressure bounds are affine bounds on π when pressures are nonnegative.

A relaxation can pass optimization and fail physics

For a normalized pipe, take $k = 1$, $q = 2$, and $\pi_i - \pi_j = 9$.

$$q^2 = 4 \leq 9 \quad \text{passes the relaxation, but} \quad 4 \neq 9.$$

For a minimization problem:

$$f^*_{\text{relaxed}} \leq f^*_{\text{original}} \leq f_{\text{feasible plan}}.$$

- A relaxation supplies a lower bound.
- A plan satisfying the original equations supplies an upper bound.
- Check every original pressure-drop equality before using the plan.
- A restriction, such as fixing pump modes, searches a smaller feasible set.

Small optimization residuals and small physical-model residuals are separate checks.

Choose the class from the equations

Model	Class / qualification
Min-cost water deliveries, linear capacities	LP
Quadratic dispatch, linear balances and bounds	Convex QP
Linear dispatch cost, convex quadratic emission cap	Convex QCQP
Hub location with Euclidean distances	SOCP after epigraph reformulation
Ellipsoidal robust feeder-capacity limits	SOCP
Pump on/off modes or exact gas pressure-drop equalities	Generally nonconvex

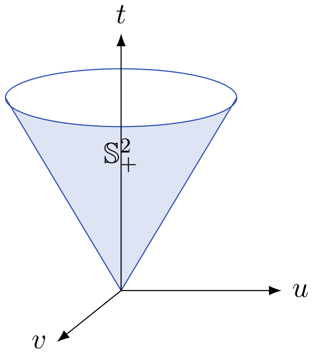
The application name does not determine the optimization class.

Two more convex families: GP and SDP

- **Geometric programs (GPs):** positive variables; minimize a posynomial, with posynomial ≤ 1 and monomial $= 1$ constraints. A monomial is $c\prod_j x_j^{a_j}$, $c > 0$; a posynomial is a sum of monomials.
- With $y_j = \log x_j$, GP becomes convex through log-sum-exp expressions. The original variables need not give a convex formulation.
- **Semidefinite programs (SDPs):** linear objectives and affine constraints with a matrix inequality $F_0 + \sum_i x_i F_i \succeq 0$.

$$X = \begin{bmatrix} t+u & v \\ v & t-u \end{bmatrix} \succeq 0$$

$$\iff \quad t \geq \sqrt{u^2 + v^2}.$$



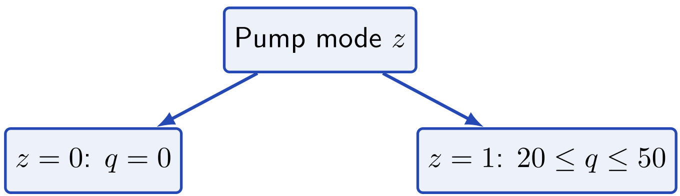
SDPs are convex, but large matrix variables can be costly. This 2-by-2 cone is SOC-representable; larger PSD cones are richer.

MOSEK: GP and semidefinite modeling.

Mixed-integer models: choose equipment and operation

$$\min_{q,z}\ cq + Fz \quad \text{s.t.} \quad 20z \le q \le 50z, \quad q \ge d, \quad z \in \{0,1\}.$$

- **MILP:** linear costs and constraints; z starts a pump and incurs fixed cost F.
- **MIQP / MISOCP:** add a convex quadratic cost or conic equipment limit. Fixing the integer choices leaves a convex subproblem.
- **Mixed-integer nonconvex NLP:** include exact nonlinear gas/water physics or nonconvex continuous costs; fixing integers does not remove all nonconvexity.



Branch-and-bound uses continuous relaxation bounds to avoid checking every combination.

With $d > 0$, the off branch is infeasible. Integer restrictions make the overall feasible set nonconvex.

Common relaxations: enlarge the set, obtain a bound

- **Continuous relaxation:** replace $z \in \{0,1\}$ by $0 \le z \le 1$. In $q \le 50z$, demand $q = 10$ can use $z = 0.2$, paying only $0.2F$.
- **Conic relaxation:** replace $q^2 = k\Delta\pi$ by $q^2 \le k\Delta\pi$. This drops part of the gas physics and gives an SOC constraint.
- **McCormick relaxation:** for $w = xy$ with $x, y \in [0,1]$, use

$$w \ge 0, \quad w \ge x + y - 1, \quad w \le x, \quad w \le y.$$

  At $x = y = 1/2$, these allow $0 \le w \le 1/2$; exact physics requires $w = 1/4$.
- **SDP lifting:** write $X = xx^T$, then relax it to $X \succeq xx^T$ using $\begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0$, retaining the other lifted constraints.

For minimization: relaxed optimum $\le$ true optimum $\le$ cost of an original-feasible plan.

A linear approximation is not automatically a relaxation: its feasible set must contain the original set.

Unconstrained optimality: use the right implication

For twice continuously differentiable f at an interior point x^*:

- **Necessary for a local minimum:** $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succeq 0$.
- **Sufficient for a strict local minimum:** $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$.
- **For convex differentiable f:** $\nabla f(x^*) = 0$ is sufficient for global optimality.

$$f(x) = -x^4 : \qquad f'(0) = 0, \quad f''(0) = 0, \quad \text{but } 0 \text{ is a maximum.}$$

A semidefinite Hessian at one point is not a sufficient minimum test.

At a capacity limit, a nonzero gradient can be optimal

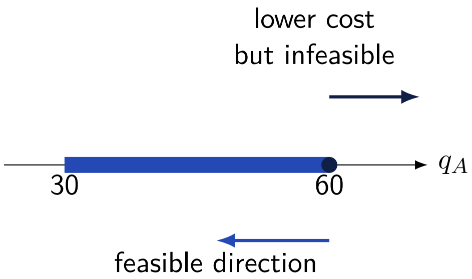
The water LP reduces to

$$\min_{30 \le q_A \le 60} 2800 - 15 q_A.$$

At $q_A^* = 60$, the derivative is -15, not zero. For differentiable convex f on a convex feasible set C, optimality is equivalent to

$$\nabla f(x^*)^T (x - x^*) \ge 0 \quad \forall x \in C.$$

There is no feasible first-order descent direction.



KKT: four pieces of an optimality certificate

For $\min f(x)$ subject to $g_i(x) \leq 0$ and $Ax = b$, define

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \sum_i \lambda_i g_i(x) + \nu^T (Ax - b).$$

Condition	Equation
Stationarity	$\nabla f(x^*) + \sum_i \lambda_i^* \nabla g_i(x^*) + A^T \nu^* = 0$
Primal feasibility	$g_i(x^*) \leq 0, \quad Ax^* = b$
Dual feasibility	$\lambda_i^* \geq 0$
Complementary slackness	$\lambda_i^* g_i(x^*) = 0$ for each i

A nonbinding inequality has zero multiplier; a binding one may also have zero multiplier.

When do KKT conditions certify the answer?

- For differentiable **convex** problems with affine equalities, a feasible KKT point is globally optimal.
- Under a suitable constraint qualification, KKT conditions are also necessary at a local optimum.
- A useful convex condition is **Slater's condition**: an appropriate relative-interior point satisfies the equalities and strictly satisfies the convex inequalities.
- For **nonconvex** problems, KKT points are generally only candidates. Multipliers need not be unique.

State convexity and regularity assumptions before claiming a certificate.

Reference: Boyd and Vandenberghe, *Convex Optimization*, Chapter 5.

KKT in dispatch: a bottleneck has a price

With $p_1 \le 160$ and demand 250, the optimum is $(160, 90)$. Use the balance sign $250 - p_1 - p_2 = 0$:

$$\mathcal{L} = C_1(p_1) + C_2(p_2) + \lambda(250 - p_1 - p_2) + \mu(p_1 - 160).$$

Other bounds are inactive, so their multipliers vanish.

$$\begin{aligned} 28 - \lambda + \mu &= 0, \\ 34 - \lambda &= 0 \end{aligned} \quad \Rightarrow \quad \lambda = 34, \quad \mu = 6.$$

- λ: local extra cost per MW of demand, in dollars/hour per MW.
- μ: local cost reduction per MW of additional Generator 1 capacity.

Complementarity: $\mu(p_1 - 160) = 6 \cdot 0 = 0$.

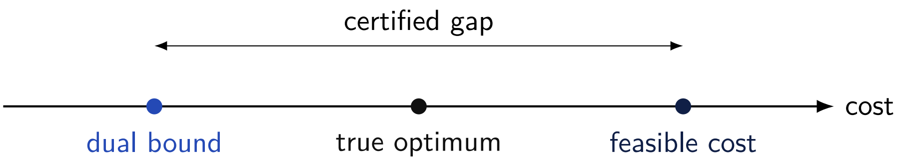
Duality turns a good plan into a bounded claim

For $\lambda \geq 0$, define the dual function

$$d(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu).$$

It gives a lower bound for the primal minimization problem:

$$d(\lambda, \nu) \leq f^* \leq f(\hat{x}) \qquad (\hat{x} \text{ primal feasible}).$$

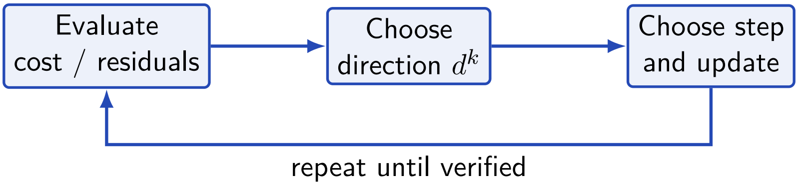


A small primal–dual gap certifies objective quality when the reported bounds are valid.

Algorithms: direction, step size, stopping rule

An iterative method constructs $x^0, x^1, \ldots$:

$$x^{k+1} = x^k + \alpha_k d^k.$$



- We use gradient and Newton methods to understand the mechanics.
- LP/QP/SOCP solvers use methods that also handle their constraints.
- Compare computation time and certificate quality, not iteration count alone.

A usable unconstrained optimization toolkit

Method	Direction / update	When to reach for it
Gradient descent	$d^k = -\nabla f(x^k)$	Smooth objectives; cheap iterations
Damped Newton	$\nabla^2 f(x^k) d^k = -\nabla f(x^k)$	Moderate size; reliable curvature
BFGS / L-BFGS	$d^k = -H_k \nabla f(x^k)$; update a curvature approximation	Smooth problems when exact Hessians are costly
Mini-batch SGD	$x^{k+1} = x^k - \alpha_k g_k$	Losses summed over many training samples

- Full-gradient methods can use line searches; stochastic methods usually use a learning-rate schedule.
- Scale variables, supply derivatives, and return both the point and termination status.

Choose the method from derivative cost, curvature, and data size.

BFGS approximates inverse curvature; L-BFGS stores a short history of vector pairs.

Algorithm 1: gradient descent with backtracking

Inputs: differentiable f, gradient, x^0, tolerance ε, budget K; $c = 10^{-4}$, $\beta = 1/2$.

1. For $k = 0, \ldots, K-1$, evaluate $g_k = \nabla f(x^k)$.
2. If $\|g_k\|_2 \le \varepsilon$, return x^k with status `stationary`.
3. Set $d_k = -g_k$ and $\alpha = 1$. While the trial point is outside the domain or

$$f(x^k + \alpha d_k) > f(x^k) + c\alpha g_k^T d_k,$$

   shrink $\alpha \leftarrow \beta\alpha$.
4. Set $x^{k+1} = x^k + \alpha d_k$ and continue.

- If the budget is exhausted, return `iteration limit`; if line search cannot make numerical progress, return `stalled`.
- For convex f, stationarity certifies a global optimum in exact arithmetic. For nonconvex f, investigate the stationary point.

This is a complete loop: compute a direction, choose a step, update, and report why it stopped.

Gradient descent on the dispatch example

For the reduced dispatch cost,

$$f(p) = 0.075(p-200)^2 + 6375, \qquad f'(p) = 0.15(p-200).$$

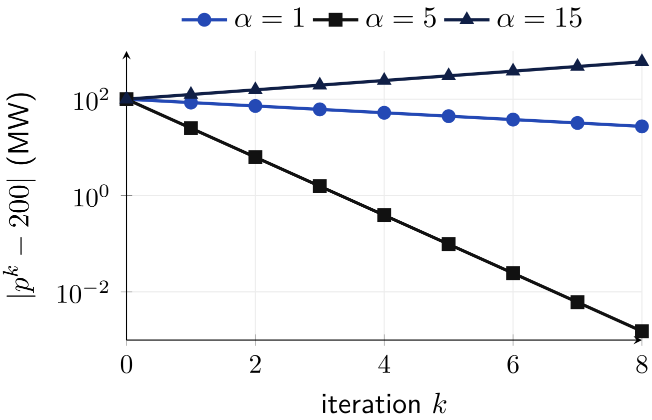
With step size $\alpha = 5$:

$$p^{k+1} = p^k - 5f'(p^k), \qquad p^{k+1} - 200 = 0.25(p^k - 200).$$

Iteration k	p_1^k (MW)	$p_2^k = 250 - p_1^k$	Cost ($/h)
0	100.0000	150.0000	7125.0000
1	175.0000	75.0000	6421.8750
2	193.7500	56.2500	6377.9297
3	198.4375	51.5625	6375.1831

For this quadratic, convergence requires $|1 - 0.15\alpha| < 1$, or $0 < \alpha < 13.33\ldots$.

Step size controls speed and stability



All curves start at $p^0 = 100$.

$$|p^k - 200| = 100r^k,$$
$$r = |1 - 0.15\alpha|.$$

The divergent run eventually leaves the physical dispatch range.

A descent direction does not make every step length safe.

Exact iterates for the unconstrained reduced quadratic; bounds require a constrained method.

Backtracking: accept sufficient decrease

For a descent direction d with $\nabla f(x)^T d < 0$:

1. Start with a trial step $\alpha = 1$.
2. Accept if the trial point is in the domain and
$$f(x + \alpha d) \le f(x) + c\alpha \nabla f(x)^T d.$$
3. Otherwise replace $\alpha \leftarrow \beta\alpha$ and repeat.

Use, for example, $c = 10^{-4}$ and $\beta = 1/2$.

For smooth f and a descent direction, sufficiently small positive steps pass this test.

This unconstrained line search does not by itself enforce network balances or equipment limits.

Newton's method uses curvature

Compute the direction by solving a linear system:

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k), \qquad x^{k+1} = x^k + \alpha_k d^k.$$

For dispatch, $f''(p) = 0.15$:

$$d = -\frac{0.15(p-200)}{0.15} = 200 - p.$$

A full step reaches 200 in one iteration for this unconstrained quadratic.

- Positive definite Hessian gives a descent direction away from stationarity.
- Damping or Hessian modification may be needed for general nonlinear costs.
- Solving the system can be expensive; exploit sparsity and avoid explicit inversion.

Newton trades more work per step for better use of local curvature.

Algorithm 2: damped Newton with a descent safeguard

Inputs: f, gradient, Hessian, x^0, tolerance ε, iteration budget.

1. Evaluate $g_k = \nabla f(x^k)$. Stop if $\|g_k\|_2 \le \varepsilon$.
2. Form $B_k = \nabla^2 f(x^k)$. If needed, add τI so $B_k \succ 0$; increase τ until a Cholesky factorization succeeds.
3. Solve $B_k d_k = -g_k$. If the direction is not numerically a descent direction, use $d_k = -g_k$.
4. Apply the same Armijo backtracking rule, starting with $\alpha = 1$; update $x^{k+1} = x^k + \alpha d_k$.
5. Return the final gradient norm and a status: stationary, stalled, or budget exhausted.

- Positive definite B_k gives $g_k^T d_k = -d_k^T B_k d_k < 0$ for $g_k \neq 0$.
- For large models, L-BFGS avoids forming the Hessian; use a tested implementation with its line search.

Near a regular local minimum, unmodified Newton steps can become fast; far away, the safeguards matter.

Constraints change both the step and the stopping test

Projected gradient for a closed convex feasible set C:

$$x^{k+1} = \Pi_C\left(x^k - \alpha \nabla f(x^k)\right), \qquad \Pi_C(y) = \arg\min_{z \in C} \|z - y\|_2^2 .$$

For reduced dispatch with $0 \le p_1 \le 160$:

$$p_1^{k+1} = \min\{160, \max\{0, p_1^k - \alpha f'(p_1^k)\}\}.$$

At $p_1 = 160$, $f'(160) = -6$, but the projected step is zero.

- Clipping handles a box; it generally does not preserve coupled balances.
- Projection onto a network-feasible set may itself require a QP.

For constrained problems, check a projected-gradient or KKT residual, not just $\|\nabla f\|$.

Convergence claims need assumptions

Method / setting	Assumptions	What follows
Gradient descent	Convex f, L-Lipschitz gradient, optimizer exists, $\alpha = 1/L$	Objective gap $O(1/k)$
Gradient descent	Also μ-strongly convex	Geometric objective-gap bound $(1-\mu/L)^k$
Newton, full steps	Near optimum; Hessian positive definite and locally Lipschitz	Local quadratic convergence of iterates

- For nonconvex f bounded below with a Lipschitz gradient, suitable steps drive gradient norms toward zero.
- Stationary points may not be minima.

Geometric convergence does not certify global optimality.

Stop when the answer is accurate enough for the system

1. **Feasibility:** balance errors and limit violations are below chosen tolerances in physical units.
2. **Optimality:** scaled KKT residuals, projected-gradient residual, or a valid primal–dual gap is small.
3. **Progress and budget:** monitor objective change, time, and iteration limits.

$$\text{Example relative gap} = \frac{f_{\text{primal}} - f_{\text{dual}}}{\max\{1, |f_{\text{primal}}|, |f_{\text{dual}}|\}}.$$

Tiny steps alone may mean stagnation. An iteration limit is not an optimality certificate.

Scale variables and residuals deliberately; a tolerance of 10^{-4} has different meanings for MW, meters, and normalized pressure.

Machine learning is often unconstrained optimization

$$\min_{\theta \in \mathbb{R}^p} \quad F(\theta) = \frac{1}{N}\sum_{i=1}^{N} \ell(\hat{y}_i(\theta), y_i) + \frac{\lambda}{2}\|\theta\|_2^2, \qquad \lambda \geq 0.$$

- **Variables:** model weights θ, rather than pump flows or generator outputs.
- **Data:** historical features and labels; the training loss measures prediction errors.
- **Regularization:** penalizes large weights without an explicit constraint.
- **Three examples:** ridge regression for demand, logistic regression for failure risk, and neural networks for nonlinear forecasting.

The optimizer learns model parameters; the learned model can later inform an infrastructure decision.

Train on past data and evaluate on held-out data; low training loss alone does not demonstrate useful forecasts.

Learning example 1: ridge regression forecasts demand

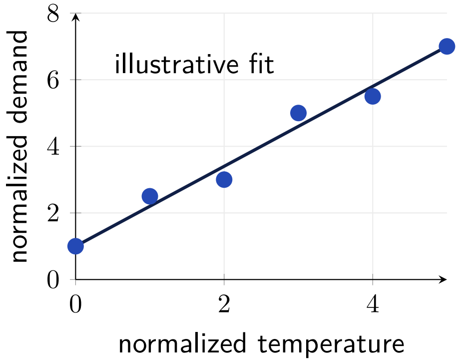
- Features a_i: temperature, hour, and past demand; target y_i: next-hour water demand.
- A linear prediction is $\hat{y}_i = a_i^T \theta$:

$$\min_{\theta} \ \frac{1}{2N} \|A\theta - y\|_2^2 + \frac{\lambda}{2} \|\theta\|_2^2 .$$

- This is an unconstrained convex QP:

$$\nabla F = \frac{1}{N} A^T (A\theta - y) + \lambda\theta.$$

- Solve a linear system or apply GD; for large N, sample rows for SGD.



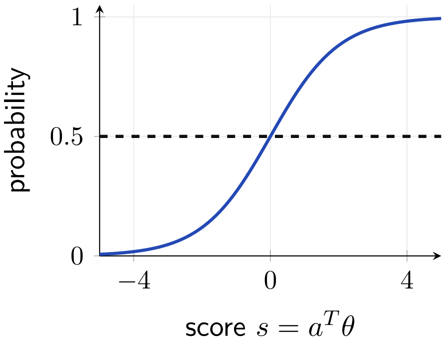
Plot illustrates a linear predictor, not a calibrated demand model. For $\lambda > 0$, this fully regularized objective is strongly convex.

Learning example 2: logistic regression predicts failures

- Label $y_i \in \{0, 1\}$: whether a pipe fails. Features a_i: age, pressure, and past leaks.
- Predict probability $p_i = \sigma(a_i^T \theta)$, with $\sigma(s) = 1/(1 + e^{-s})$.
- Fit the unconstrained convex loss:

$$\min_{\theta} \quad \frac{1}{N} \sum_i \left[\log(1 + e^{a_i^T \theta}) - y_i a_i^T \theta\right] + \frac{\lambda}{2} \|\theta\|_2^2 .$$

- Each sample contributes gradient $(p_i - y_i) a_i$; average for GD, sample for SGD.

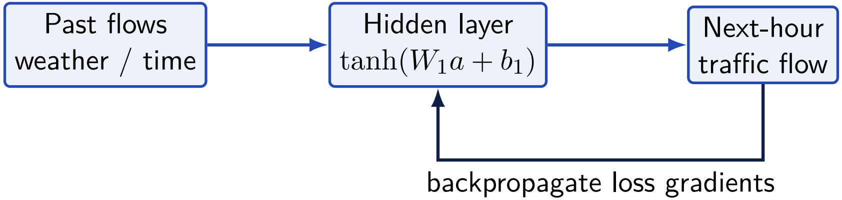


A nonlinear prediction function can still lead to a convex training problem.

Boyd and Vandenberghe: statistical estimation, logistic regression. Use stable log-sum-exp evaluation in code.

Learning example 3: a neural network forecasts traffic

$$\hat{y}_i = W_2 \tanh(W_1 a_i + b_1) + b_2, \qquad \min_{\theta} \ \frac{1}{2N} \sum_i \|\hat{y}_i - y_i\|_2^2 + \frac{\lambda}{2} \|\theta\|_2^2 .$$



- All weights and biases are unconstrained real decision variables.
- The loss is generally **nonconvex** in the joint parameter vector θ.
- Backpropagation computes gradients; mini-batch SGD uses them to update the weights. Initialization and learning rate affect the result.

The update rule still works with gradients, but it no longer carries a global-optimum guarantee.

Algorithm 3: mini-batch stochastic gradient descent

Inputs: training samples, initial θ^0, batch size b, learning rates α_k, update budget K.

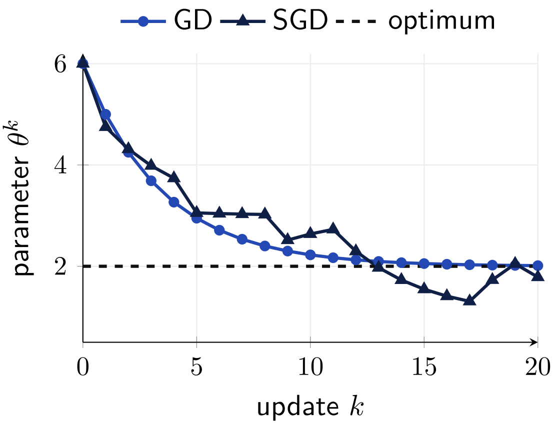
1. Sample a batch $\mathcal{B}_k$ of b indices uniformly from $\{1, \ldots, N\}$.
2. Compute a noisy estimate of the full gradient:
$$g_k = \frac{1}{b} \sum_{i \in \mathcal{B}_k} \nabla_\theta \ell_i(\theta^k) + \lambda \theta^k.$$
3. Update $\theta^{k+1} = \theta^k - \alpha_k g_k$.
4. Periodically measure full training loss and validation loss. Save the best validation checkpoint; stop at the chosen budget or validation rule.

- Uniform sampling makes g_k unbiased; larger batches usually reduce its variance at greater cost.
- A constant learning rate can leave a noise floor. A decreasing schedule can reduce noise effects under suitable assumptions.

A small batch gradient is not a stationarity certificate for the full training objective.

Optimization Methods for Supervised Machine Learning (2017). Validation stopping targets generalization, not an optimality certificate.

GD versus SGD: smooth progress versus noisy updates



- Two samples: $y_1 = 1$, $y_2 = 3$; $F(\theta) = \frac{1}{4}\sum_{i=1}^{2}(\theta - y_i)^2$.
- $\nabla F = \theta - 2$, so $\theta^* = 2$.
- Both start at 6, with $\alpha = 0.25$. SGD samples one y_i per update.
- GD averages both gradients; SGD fluctuates because individual targets differ.

Compare work as well as updates: one GD step uses all N samples; one SGD step uses only its batch.

Exact toy-problem iterates with a fixed random seed. Constant-step SGD illustrates a noise floor, not monotone descent.

From a mathematical QP to executable code

```
import cvxpy as cp

p = cp.Variable(2)
cost = (0.025 * cp.square(p[0])
        + 20 * p[0]
        + 0.05 * cp.square(p[1])
        + 25 * p[1])
constraints = [cp.sum(p) == 250,
               p >= 0, p <= 300]
problem = cp.Problem(cp.Minimize(cost),
                     constraints)
problem.solve()
print(problem.status)
print(p.value, problem.value)
```

Modeling layer
Expresses variables, objectives, and constraints; converts the model into solver form.

Numerical solver
Computes a candidate solution and reports status and numerical diagnostics.

Expected result: approximately $(200, 50)$ and 6375.

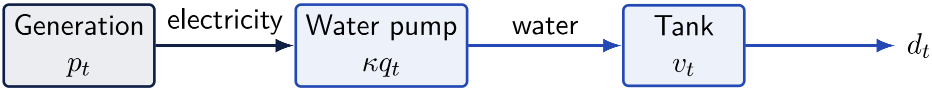
Requires CVXPY and a compatible installed solver. Official CVXPY QP documentation.

Validate the engineering answer after the solve

Check	Questions to ask
Balances	Do generation, withdrawals, storage changes, and flows balance?
Operating limits	Are capacities, pressure bounds, and queue limits respected?
Original physics	Does a relaxed or linearized solution satisfy the original equations closely enough?
Sensitivity	Does the plan remain acceptable when demand or a component changes?
Decision meaning	Are modes implementable? Are units and time intervals consistent?

The optimizer solves the model you wrote. Validation connects that model to the system.

Studio exercise: coordinate power and water



$$\begin{aligned} \min \quad & \sum_t (a p_t^2 + b p_t) \\ \text{s.t.} \quad & p_t = \ell_t + \kappa q_t, \\ & v_{t+1} = v_t + \Delta t (q_t - d_t), \\ & 0 \le p_t \le \bar{p}, \quad 0 \le q_t \le \bar{q}, \quad 0 \le v_t \le V. \end{aligned}$$

Given initial storage, impose $v_{T+1} = v_1$ and bound all storage states.

Classify this model. Then add an inverter capability limit or pump on/off mode.

$a \ge 0$; ℓ_t is non-pump electrical load. Fixed head/efficiency makes κ constant.

Studio debrief: one change can change the problem class

Modification	Resulting model
Quadratic generation cost, fixed κ, affine limits	Convex QP
Linear generation cost instead	LP
Add $p_t^2 + Q_t^2 \le S^2$ with fixed S	Convex QCQP; conic-representable
Add ellipsoidal uncertain-load capacity constraints	SOCP-representable convex model
Add binary pump modes	Mixed-integer model
Replace fixed κ with exact variable-head power physics	Generally nonlinear and nonconvex

Five questions to take into the next module

1. What are the decisions, units, objective, and physical constraints?
2. Can I mix feasible operating plans and stay feasible?
3. Is the model an LP, QP, QCQP, SOCP, or a nonconvex variant?
4. What certifies optimality, and which bottleneck has value?
5. Does the computed answer satisfy the original system model?

Good infrastructure optimization combines a useful model, a suitable method, and a checked answer.

Reading and resources

- Boyd and Vandenberghe, *Convex Optimization*. Chapters 2–5 and 9.
  https://web.stanford.edu/~boyd/cvxbook/
- Nocedal and Wright, *Numerical Optimization*, 2nd ed. Chapters 2–3 and 12.
- MOSEK Modeling Cookbook, Chapter 3: conic quadratic modeling.
  docs.mosek.com/modeling-cookbook/cqo.html
- CVXPY: worked quadratic and second-order cone examples.
  Quadratic program | Second-order cone program

Questions, discussions, and ideas

Which infrastructure decision would you optimize?

Choose a system, identify one controllable quantity, and name one constraint you cannot violate.

Dr. Ahmad F. Taha

https://ahmadstatespace.github.io/

ahmad.taha@vanderbilt.edu