

These instructor solutions show the intermediate calculations as well as the final results. Numerical values and figures were produced by the reference Python implementation in code/homework3_fa26_python.py.

Problem 1: Economic dispatch

Solution. Use the balance constraint to eliminate $P_{G_2} = P_L - P_{G_1}$. Apart from the constant π , the resulting scalar objective is

$$\phi(P_{G_1}) = 0.025P_{G_1}^2 + 20P_{G_1} + 0.05(P_L - P_{G_1})^2 + 25(P_L - P_{G_1}).$$

Its derivatives are

$$\phi'(P_{G_1}) = 0.15P_{G_1} - 0.10P_L - 5, \quad \phi''(P_{G_1}) = 0.15 > 0.$$

Therefore the stationary point is the unique minimum whenever both outputs are nonnegative:

$$P_{G_1}^* = \frac{2}{3}P_L + \frac{100}{3}, \quad P_{G_2}^* = \frac{1}{3}P_L - \frac{100}{3}.$$

For $P_L = 250$ MW this gives $(P_{G_1}^*, P_{G_2}^*) = (200, 50)$ MW and

$$f^* = 0.025(200)^2 + 20(200) + 0.05(50)^2 + 25(50) + \pi = 6375 + \pi \approx 6378.1416.$$

The six optimal pairs, in increasing order of load, are

$$(133.333, 16.667), (166.667, 33.333), (200, 50), (233.333, 66.667), (266.667, 83.333), (300, 100).$$

Equal marginal costs explain this split. Increasing a unit's quadratic coefficient makes its marginal cost rise faster, shifting incremental generation toward the other unit.

At $P_L = 250$ the unconstrained value $P_{G_1} = 200$ violates the new cap. Thus the cap binds and balance forces $(P_{G_1}, P_{G_2}) = (160, 90)$ MW. The value is $6495 + \pi \approx 6498.1416$. Finally, adding π raises every feasible objective by the same amount: it changes the optimal value but cannot change the optimizer.

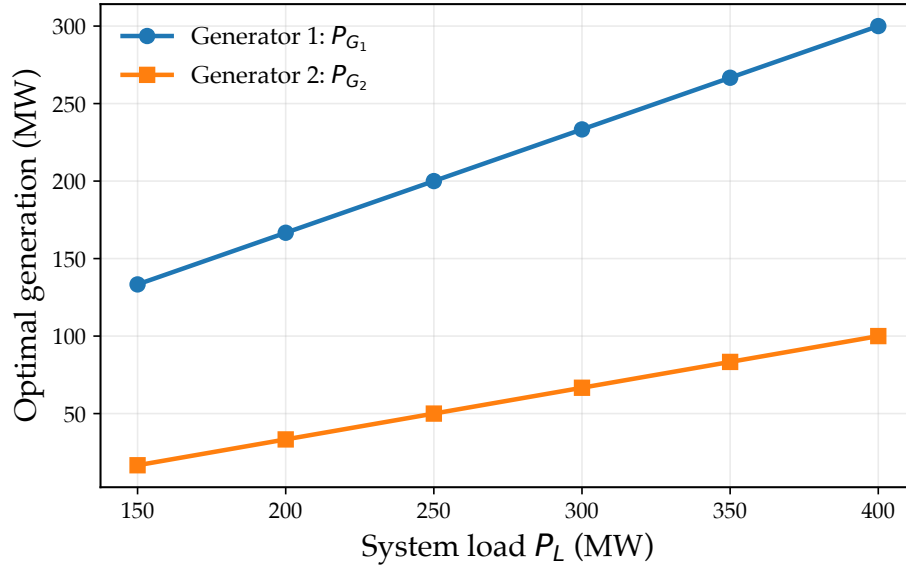


Figure 1: Optimal sharing of demand by the two generators.

Problem 2: A halfspace is a convex cone

Solution. A set K is a convex cone if $x, y \in K$ and $\lambda, \mu \geq 0$ imply $\lambda x + \mu y \in K$. Take $x, y \in \mathcal{S}$. Since $a^T x \leq 0$ and $a^T y \leq 0$,

$$a^T(\lambda x + \mu y) = \lambda a^T x + \mu a^T y \leq 0$$

for all nonnegative λ and μ . Therefore $\lambda x + \mu y \in \mathcal{S}$, proving the cone property and convexity simultaneously. Negative multipliers are not part of the cone definition; allowing them could reverse the inequality.

Problem 3: Convexity of a slab

Solution. Let $x, y \in \mathcal{S}$ and $0 \leq \theta \leq 1$. Linearity gives

$$a^T(\theta x + (1 - \theta)y) = \theta a^T x + (1 - \theta)a^T y.$$

Because both scalar values lie in $[\alpha, \beta]$,

$$\alpha = \theta \alpha + (1 - \theta)\alpha \leq a^T(\theta x + (1 - \theta)y) \leq \theta \beta + (1 - \theta)\beta = \beta.$$

The full segment between x and y is therefore in the slab. Equivalently, the slab is the intersection of the two convex halfspaces $a^T x \leq \beta$ and $-a^T x \leq -\alpha$.

Problem 4: Convexity of the radius-two ball

Solution. If $x, y \in \Omega$ and $0 \leq \theta \leq 1$, the triangle inequality and positive homogeneity of the Euclidean norm give

$$\|\theta x + (1 - \theta)y\|_2 \leq \theta \|x\|_2 + (1 - \theta) \|y\|_2 \leq 2.$$

Squaring both nonnegative sides proves that the convex combination belongs to Ω . For the sublevel-

set proof, $h(y) = \|y\|_2^2$ has Hessian $2I \succeq 0$, so it is convex, and $\Omega = \{y : h(y) \leq 4\}$ is a convex sublevel set.

Problem 5: Scalar description of the PSD cone

Solution. For a real symmetric 3×3 matrix, positive semidefiniteness is equivalent to nonnegativity of every principal minor. The three order-one conditions are

$$x_1 \geq 0, \quad x_4 \geq 0, \quad x_6 \geq 0.$$

The three order-two conditions are

$$x_1x_4 - x_2^2 \geq 0, \quad x_1x_6 - x_3^2 \geq 0, \quad x_4x_6 - x_5^2 \geq 0.$$

Expansion of the order-three determinant yields

$$\det A = x_1x_4x_6 + 2x_2x_3x_5 - x_1x_5^2 - x_4x_3^2 - x_6x_2^2 \geq 0.$$

These seven polynomial inequalities are jointly necessary and sufficient. Notice that PSD does not require the off-diagonal entries x_2, x_3, x_5 themselves to be nonnegative.

Problem 6: A parameterized quadratic

Solution. Write $f(x) = x^T Q x$, where

$$Q = \begin{bmatrix} 2 & \beta & -1 \\ \beta & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix}.$$

Since $\nabla^2 f = 2Q$, the function is convex exactly when $Q \succeq 0$. The order-one minors are positive. The order-two principal minors are

$$2 - \beta^2, \quad 5, \quad 2,$$

and direct expansion gives

$$\det Q = 3 - 2\beta - 3\beta^2.$$

Thus $|\beta| \leq \sqrt{2}$ and $3\beta^2 + 2\beta - 3 \leq 0$. Solving the second inequality gives

$$\frac{-1 - \sqrt{10}}{3} \leq \beta \leq \frac{-1 + \sqrt{10}}{3}.$$

This interval is contained in $[-\sqrt{2}, \sqrt{2}]$, so it satisfies every minor condition and is both necessary and sufficient.

Problem 7: Function classification

Solution. For $f(x) = \log x$ on $\mathbb{R}_{++}$,

$$f'(x) = \frac{1}{x}, \quad f''(x) = -\frac{1}{x^2} < 0.$$

It is therefore strictly concave and not convex.

For $g(x_1, x_2) = x_1^2/x_2$ with $x_2 > 0$,

$$\nabla g = \begin{bmatrix} 2x_1/x_2 \\ -x_1^2/x_2^2 \end{bmatrix}, \quad \nabla^2 g = \frac{2}{x_2^3} \begin{bmatrix} x_2^2 & -x_1x_2 \\ -x_1x_2 & x_1^2 \end{bmatrix} = \frac{2}{x_2^3} \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix}^\top \succeq 0.$$

Hence g is convex. The Hessian has determinant zero, so this calculation does not establish strict convexity; indeed g is linear along positive rays. It is also the perspective of the convex scalar quadratic $u \mapsto u^2$.

Problem 8: Five objectives on a polyhedron

Solution. The two boundary lines intersect at $(4/5, 3/5)$. Together with the axes, the lower boundary has vertices $(0, 3)$, $(4/5, 3/5)$, and $(2, 0)$, while the feasible set extends to the northeast.

Part	Optimal set	Optimal value
(a)	$\{(4/5, 3/5)\}$	$7/5$
(b)	none; unbounded below	$-\infty$
(c)	$\{(x_1, 0) : x_1 \geq 2\}$	0
(d)	$\{(3/4, 3/4)\}$	$3/4$
(e)	$\{(9/13, 12/13)\}$	$36/13$

For (a), evaluating the linear objective at the three lower vertices gives $3, 7/5, 2$. For (b), (t, t) is feasible for sufficiently large t and has value $-2t \rightarrow -\infty$. For (c), the smallest possible x_2 is zero, and feasibility then requires $x_1 \geq 2$. For (d), minimizing $\max\{x_1, x_2\}$ balances the coordinates; setting $x_1 = x_2 = t$ gives $4t \geq 3$, hence $t = 3/4$. For (e), the optimum lies on $3x_1 + x_2 = 3$. Substitute $x_2 = 3 - 3x_1$:

$$4x_1^2 + (3 - 3x_1)^2 = 13x_1^2 - 18x_1 + 9,$$

whose derivative vanishes at $x_1 = 9/13$, giving $x_2 = 12/13$.

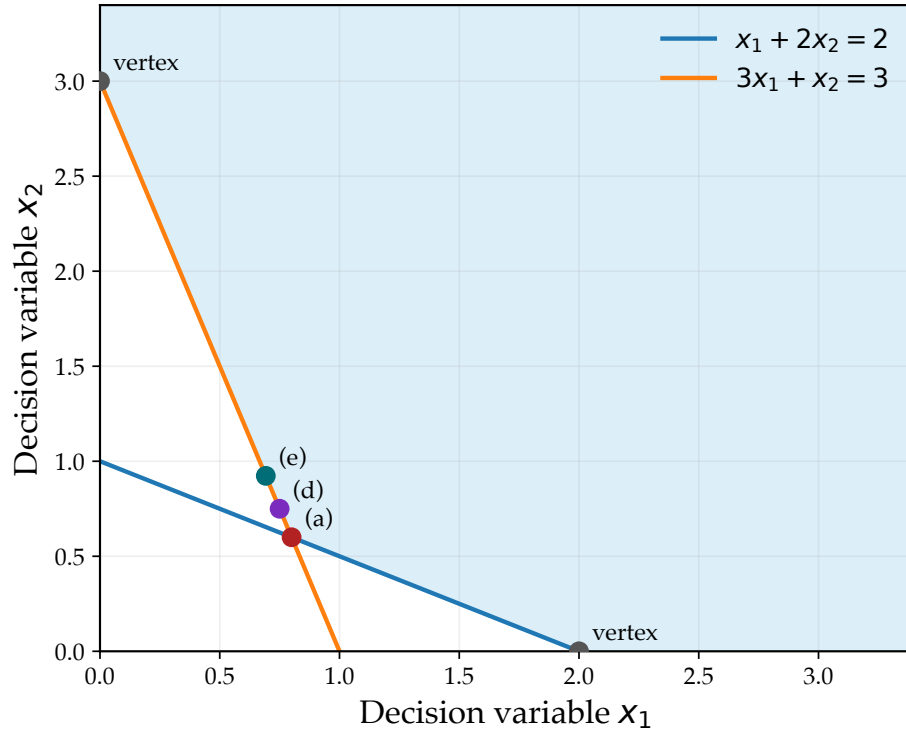


Figure 2: Feasible polyhedron and the key boundary points used in Problem 8.

Problem 9: First-order and KKT certificates on a box

Solution. All leading principal minors of P are positive: 6, 29, and 98. Thus $P \succ 0$ and the objective is strictly convex. At $x^* = (-1, 1/4, 1)^\top$,

$$\nabla f(x^*) = Px^* + q = (2, 0, -3)^\top.$$

For any feasible x ,

$$\nabla f(x^*)^\top (x - x^*) = 2(x_1 + 1) - 3(x_3 - 1) \geq 0,$$

because $x_1 \geq -1$ and $x_3 \leq 1$. The first-order variational inequality proves global optimality, and strict convexity proves uniqueness.

For a KKT certificate, write the lower bounds as $-x_i - 1 \leq 0$ with multipliers $\lambda_i^- \geq 0$, and the upper bounds as $x_i - 1 \leq 0$ with multipliers $\lambda_i^+ \geq 0$. Stationarity is

$$Px^* + q - \lambda^- + \lambda^+ = 0.$$

Choose $\lambda^- = (2, 0, 0)^\top$ and $\lambda^+ = (0, 0, 3)^\top$. They are dual feasible; their associated constraints are active; and every inactive constraint has zero multiplier. Hence primal feasibility, stationarity, dual feasibility, and complementary slackness all hold. Substitution gives $f(x^*) = -8.15625$.

Problem 10: KKT with two inequalities

Solution. Put both constraints in the form

$$g_1(x) = 3.5 - x_1 - x_2 \leq 0, \quad g_2(x) = 1 + \frac{1}{2}x_1^2 - x_2 \leq 0.$$

The Lagrangian is $L = x_1^2 + 2x_2^2 + \lambda_1 g_1 + \lambda_2 g_2$. Its stationarity equations are

$$2x_1 - \lambda_1 + \lambda_2 x_1 = 0, \quad 4x_2 - \lambda_1 - \lambda_2 = 0.$$

There are four active-set cases. With neither active, $x = (0, 0)$ is infeasible. With only g_1 active, stationarity and $x_1 + x_2 = 3.5$ give $(7/3, 7/6)$, which violates $g_2 \leq 0$. With only g_2 active, the stationary candidate is $(0, 1)$, which violates $g_1 \leq 0$.

When both are active, eliminate x_2 from $x_2 = 3.5 - x_1 = 1 + x_1^2/2$. This yields

$$x_1 = -1 \pm \sqrt{6}, \quad x_2 = 4.5 \mp \sqrt{6}.$$

Solving the stationarity equations at the positive-root point gives

$$x^* = (\sqrt{6} - 1, 4.5 - \sqrt{6}) \approx (1.44949, 2.05051),$$

$$\lambda_1 \approx 6.03708, \quad \lambda_2 \approx 2.16497.$$

Both multipliers are nonnegative. The negative-root intersection gives a negative inequality multiplier and is rejected. The objective value at x^* is approximately 10.51021.

The objective is strictly convex; g_1 is affine and g_2 is convex. The point $(1.5, 3)$ is strictly feasible, so Slater's condition holds. Consequently the valid KKT point is the unique global minimizer and therefore also a strict local minimizer.

Problem 11: KKT on a line and ellipse

Solution. Let $h = x_1 - x_2 - 4 = 0$ and $g = 4x_1^2 + x_2^2 - 25 \leq 0$. If the ellipse were inactive, minimizing on the line $x_1 = x_2 + 4$ would give $(3, -1)$, but $4(3)^2 + (-1)^2 = 37 > 25$, so the ellipse must be active.

Substituting $x_2 = x_1 - 4$ into the ellipse boundary gives $5x_1^2 - 8x_1 - 9 = 0$. Thus the two feasible endpoints are

$$(x_1, x_2) = \left(\frac{4 \pm \sqrt{61}}{5}, \frac{-16 \pm \sqrt{61}}{5} \right).$$

For

$$L = 2x_1 + x_2^2 + \nu h + \lambda g,$$

stationarity requires

$$2 + \nu + 8\lambda x_1 = 0, \quad 2x_2 - \nu + 2\lambda x_2 = 0.$$

At the plus-root point, $\lambda \approx 0.081681$ and $\nu \approx -3.54348$, so dual feasibility holds. At the minus-root point, $\lambda \approx -0.481681$, which violates $\lambda \geq 0$. Therefore

$$x^* = \left(\frac{4 + \sqrt{61}}{5}, \frac{-16 + \sqrt{61}}{5} \right) \approx (2.36205, -1.63795), \quad f^* \approx 7.40698.$$

The objective and inequality constraint are convex, the equality is affine, and a relative-interior feasible point exists on the line. The valid KKT certificate is therefore sufficient for global optimality.

Problem 12: Newton steps and the role of α

Solution. The derivatives and minimizer are

$$\nabla f(x) = \begin{bmatrix} 2x_1 - 2 \\ 3x_2 + 3 \end{bmatrix}, \quad H = \nabla^2 f = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \quad x^* = (1, -1)^T.$$

Since $\nabla f(x) = H(x - x^*)$, the Newton direction is

$$d = -H^{-1}\nabla f(x) = -(x - x^*) = x^* - x.$$

Here a “full step” means $\alpha = 1$, so the entire computed direction is used; it has nothing to do with a physical time increment. From $x^{(0)} = (-1, 2)$,

$$x^{(1)} = x^{(0)} + d^{(0)} = x^* = (1, -1),$$

and every later iterate remains at x^* .

For a general constant α , let $e^{(k)} = x^{(k)} - x^*$. Then

$$e^{(k+1)} = (1 - \alpha)e^{(k)}, \quad \|e^{(k)}\|_2 = |1 - \alpha|^k \|e^{(0)}\|_2.$$

Thus convergence occurs exactly for $0 < \alpha < 2$. With $\alpha = 0.1$, the error shrinks by 0.9 each iteration. With $\alpha = 1$, it becomes zero in one step. With $\alpha = 1.5$, it alternates sign while shrinking by 0.5. With $\alpha = 5$, it grows by a factor of 4 and alternates sign; the first iterates are $(9, -13)$, $(-31, 47)$, and $(129, -193)$.

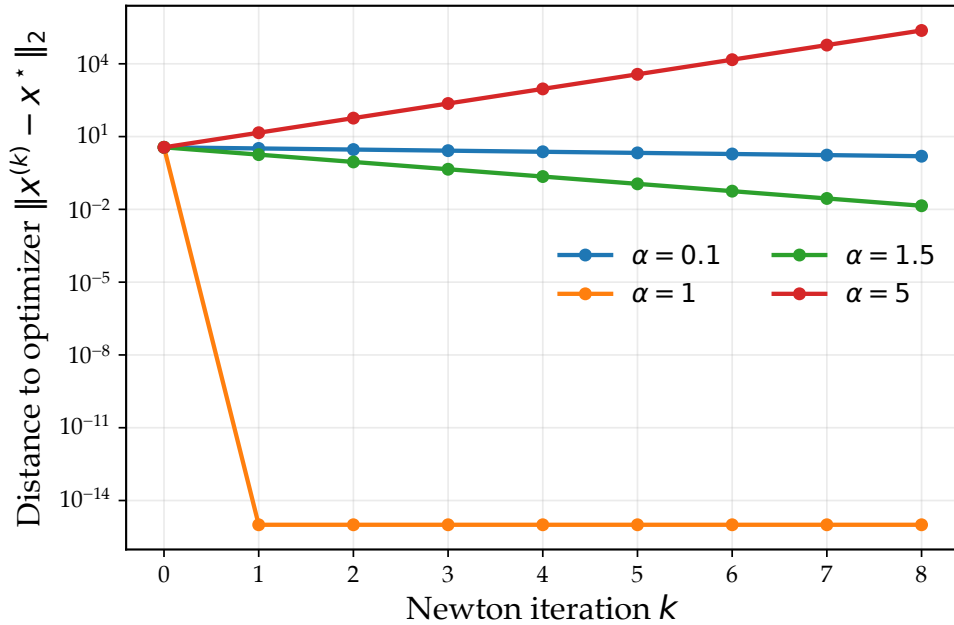


Figure 3: Newton error for several explicit choices of the step multiplier α .

Problem 13: Fixed-step gradient (steepest) descent

Solution. In the Euclidean norm, the unit direction of most rapid local decrease is $-\nabla f / \|\nabla f\|_2$; absorbing its length into the step size gives the usual gradient-descent update $x^+ = x - \alpha \nabla f(x)$.

Here

$$Q = \nabla^2 f = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix}, \quad \nabla f(x) = Qx, \quad x^* = 0.$$

Starting at $(3, -1)^\top$ with $\alpha = 1/4$,

$$g^{(0)} = (20, -8)^\top, \text{quad}x^{(1)} = (3, -1) - \frac{1}{4}g^{(0)} = (-2, 1)^\top,$$

$$g^{(1)} = (-14, 6)^\top, \text{quad}x^{(2)} = (-2, 1) - \frac{1}{4}g^{(1)} = (1.5, -0.5)^\top.$$

The error iteration is $e^{(k+1)} = (I - \alpha Q)e^{(k)}$. The eigenvalues of Q are

$$m = 4 - 2\sqrt{2}, \quad L = 4 + 2\sqrt{2}.$$

Convergence from every initial point requires $|1 - \alpha\lambda_i| < 1$ for both eigenvalues, equivalently

$$0 < \alpha < \frac{2}{L} = \frac{1}{2 + \sqrt{2}} \approx 0.292893.$$

The constant step that minimizes the worst eigen-direction contraction is

$$\alpha^* = \frac{2}{L + m} = \frac{1}{4}, \quad \rho^* = \frac{L - m}{L + m} = \frac{1}{\sqrt{2}}.$$

On the requested finite grid and after exactly 40 iterations, $\alpha = 0.24$ gives the smallest final gradient norm. That finite-horizon result does not contradict the analytic worst-case optimum $1/4$.

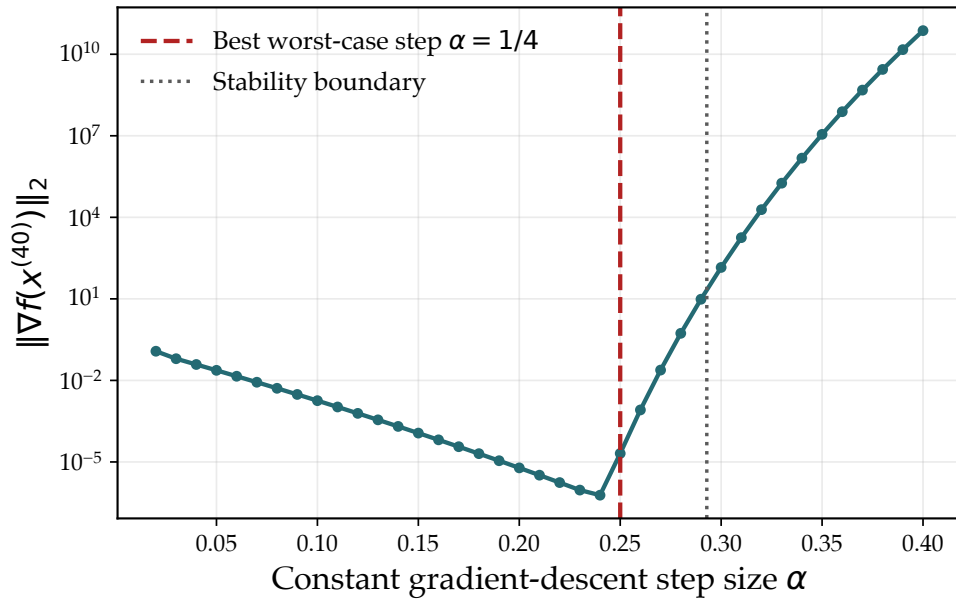


Figure 4: Final residual after 40 iterations; the vertical line marks the stability limit.

Problem 14: Gradient descent with exact line search

Solution. The gradient is

$$\nabla f(x) = \begin{bmatrix} 4(x_1 - 2)^3 \\ 4(x_2 + 1) \\ 12(x_3 - 3)^3 \end{bmatrix}.$$

At iteration k , set $d^{(k)} = -\nabla f(x^{(k)})$ and define the scalar function

$$\phi_k(\alpha) = f(x^{(k)} + \alpha d^{(k)}), \quad \alpha \geq 0.$$

An exact line search chooses a global nonnegative minimizer. At an interior minimizer it satisfies

$$\phi'_k(\alpha) = \nabla f(x^{(k)} + \alpha d^{(k)})^\top d^{(k)} = 0.$$

This is different from fixed-step gradient descent: the descent direction is the same, but a one-dimensional optimization is solved anew at every iteration.

The first three computed steps are

k	α_k	$x_1^{(k+1)}$	$x_2^{(k+1)}$	$x_3^{(k+1)}$
0	0.02325705	2	-0.0930282	2.7673228
1	0.25025489	2	-1.0009247	2.8051519
2	1.71892347	2	-0.9945668	2.9577420

The corresponding objective values at $k = 0, 1, 2, 3$ are 50, 1.6539886, 0.00432592, and 6.86065×10^{-5} . The reference implementation satisfies the successive-step tolerance after 112 iterations at approximately $(2, -1, 2.9976119)$. The first coordinate begins at its optimum and therefore never moves. Near $x_3 = 3$, the quartic has vanishing curvature, so the final progress is slow even though each line search is exact.

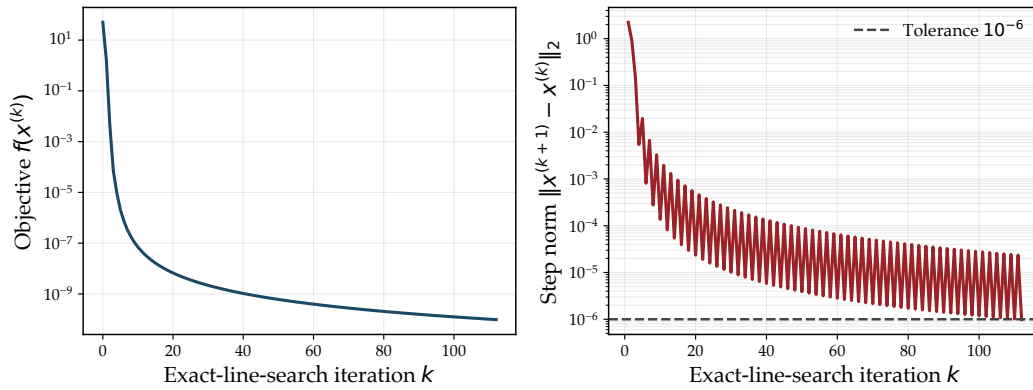


Figure 5: Objective and successive-step norm for exact-line-search gradient descent.

Problem 15: Deriving equality standard form for an LP

Solution. Start from the class formulation

$$\min_x c^\top x + c_0 \quad \text{s.t.} \quad A_{\leq} x \leq b_{\leq}, \quad A_{\geq} x \geq b_{\geq}, \quad Ex = d,$$

with any variable bounds. Each original variable is written as $x = Tz + t$ with new nonnegative

variables:

original restriction	substitution
x_j free	$x_j = x_j^+ - x_j^-, \quad x_j^+, x_j^- \geq 0$
$x_j \geq \ell_j$	$x_j = \ell_j + z_j, \quad z_j \geq 0$
$x_j \leq u_j$	$x_j = u_j - z_j, \quad z_j \geq 0$
$\ell_j \leq x_j \leq u_j$	$x_j = \ell_j + z_j, \quad z_j + s_j^u = u_j - \ell_j$

For each $\leq$ row add a slack $s \geq 0$; for each $\geq$ row subtract a surplus $s \geq 0$. After substituting $x = Tz + t$, the equality blocks are

$$A_{\leq}Tz + s_{\leq} = b_{\leq} - A_{\leq}t,$$

$$A_{\geq}Tz - s_{\geq} = b_{\geq} - A_{\geq}t, \quad ETz = d - Et.$$

The transformed objective is

$$c^T(Tz + t) + c_0 = (T^T c)^T z + (c^T t + c_0).$$

Collecting variables and rows gives $\min\{\tilde{c}^T z + \kappa : \tilde{A}z = \tilde{b}, z \geq 0\}$. A maximization is converted by negating its objective.

For the example, let $x_1 = u - v$ and $x_2 = -1 + y$ with $u, v, y \geq 0$. The upper bound $x_2 \leq 3$ becomes $y \leq 4$. A complete standard form is

$$\begin{aligned} &\text{minimize} && 2u - 2v + y - 1 \\ &\text{subject to} && u - v + 2y - s_1 = 5, \\ & && -u + v + y + s_2 = 5, \\ & && y + s_3 = 4, \\ & && u, v, y, s_1, s_2, s_3 \geq 0. \end{aligned}$$

The first row comes from the $\geq$ constraint, so it uses a subtracted surplus. The second and third rows come from $\leq$ constraints, so they use added slacks. One optimizer is

$$(u, v, y, s_1, s_2, s_3) = (0, 5/3, 10/3, 0, 0, 2/3).$$

It maps back to $(x_1, x_2) = (-5/3, 7/3)$ and has objective -1 . Direct substitution verifies both original inequalities and the bounds.

Problem 16: Regional Walmart location as an SOCP

Solution. Let $a_i \in \mathbb{R}^2$ be the coordinates in `upper_cumberland_coordinates.csv`, let p_i be population, and set $w_i = p_i / \sum_j p_j$. The epigraph variable t_i is an upper bound on distance to town i . Because every $w_i > 0$, minimizing the objective forces $t_i = \|s - a_i\|_2$ at an optimum. The variable r is an upper bound on every t_i , hence it represents the worst community distance.

A complete SOCP is

$$\begin{aligned} &\text{minimize}_{s,t,r} && \sum_{i=1}^6 w_i t_i + 0.15r \\ &\text{subject to} && \|s - a_i\|_2 \leq t_i, \quad i = 1, \dots, 6, \\ & && t_i \leq r, \quad i = 1, \dots, 6, \\ & && -8 \leq s_1 \leq 12, \quad -5 \leq s_2 \leq 15. \end{aligned}$$

Each norm inequality is a second-order-cone constraint; all remaining constraints are linear. CVXPY gives

$$s^* \approx (6.33862, 3.61840) \text{ km}, \quad r^* \approx 27.47281 \text{ km}, \quad f^* \approx 23.00256.$$

The individual distances in the table order are

$$t^* \approx (2.20174, 17.70777, 16.26557, 25.10341, 27.47280, 27.47281) \text{ km.}$$

Thus Gainesboro and Sparta jointly determine the worst-distance term to numerical tolerance. Moving the coefficient 0.15 toward zero emphasizes the population-weighted sum and approaches a weighted geometric-median location. Increasing it toward one places more emphasis on protecting the farthest community and pulls the solution toward a minimax center.

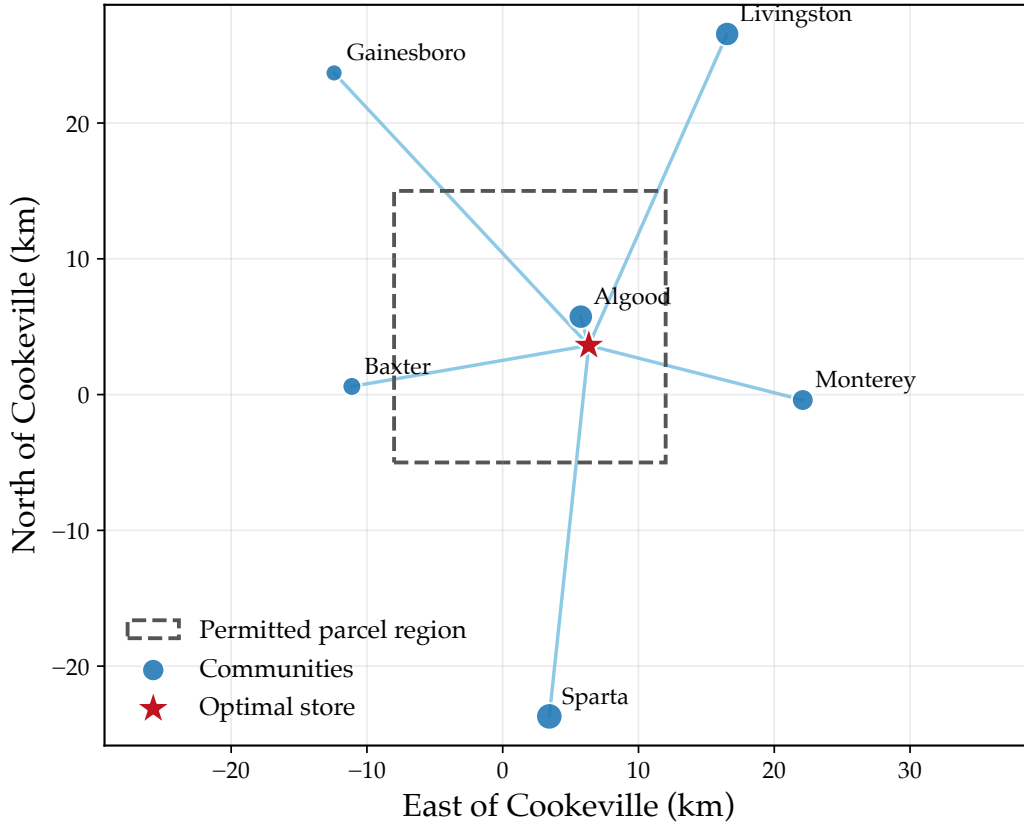


Figure 6: Town network, permitted parcel rectangle, and optimal store location. Node size represents population.

Problem 17: Comparing convex relaxations

Solution. For R1–R3, replace (x_1x_2, x_3x_4, x_2x_5) by (y_{12}, y_{34}, y_{25}) . The common relaxed objective is

$$-b^T x - 4y_{12} - 3y_{34} + 2y_{25}, \quad a^T x \leq 10, \quad 0 \leq x \leq 1.$$

R1 imposes $0 \leq y_{ij} \leq x_i$ and $y_{ij} \leq x_j$. R2 also imposes the missing lower McCormick plane $y_{ij} \geq x_i + x_j - 1$. R3 adds every listed cover cut. All three are LPs.

For R4, use X_{ij} for products and impose

$$\begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0, \quad X_{ii} = x_i,$$

together with pairwise McCormick inequalities and the cover cuts. Multiplying the nonnegative

budget slack $10 - a^\top x$ by x_j and by $1 - x_j$ gives the RLT inequalities

$$a^\top X_{:,j} \leq 10x_j, \quad 10 - a^\top x - 10x_j + a^\top X_{:,j} \geq 0, \quad j = 1, \dots, 5.$$

R4 is therefore an SDP.

Enumeration of all 32 binary vectors gives the exact solution $x = (0, 0, 1, 1, 1)$ with objective -21 . The relaxation results are

Model	Lower bound	Relaxed x
R1	-22.909091	(.0909, .0909, 1, 1, 1)
R2	-22.727273	(.0909, .0909, 1, 1, 1)
R3	-22.461538	(.1538, .1538, .9231, .9231, .9231)
R4	-21.730955	(.1614, .1737, .9062, .9062, .9099)
Exact	-21	(0, 0, 1, 1, 1)

Because this is a minimization problem, enlarging the feasible set can only lower the objective, so every relaxation gives a lower bound. Adding valid constraints removes fractional points without removing binary-feasible points. The bound consequently rises toward -21 ; R4 has the smallest gap here. Tighter bounds are useful in branch-and-bound because they can prune more nodes and certify optimality sooner.

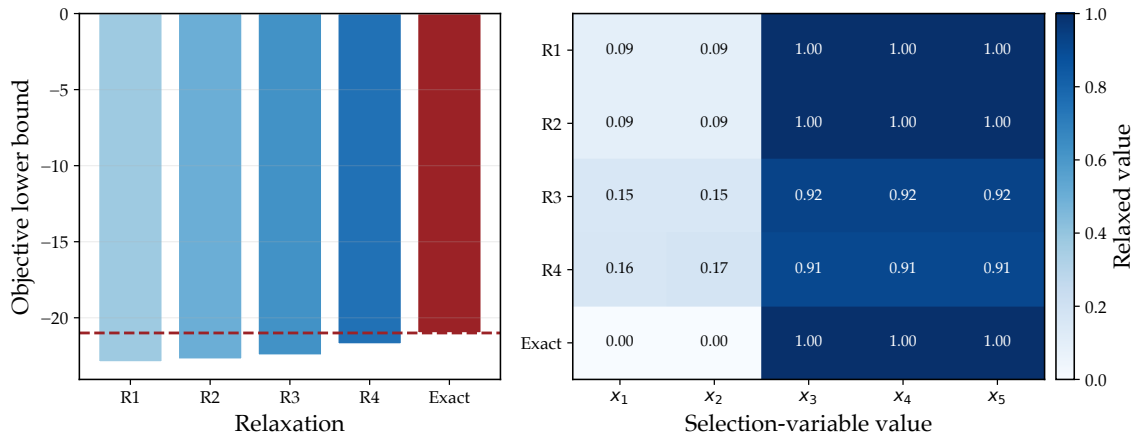


Figure 7: Relaxation bounds and fractional selection vectors compared with the exact binary optimum.

Problem 18: Bike-demand ridge regression

Solution. With design matrix A , response y , and $D = \text{diag}(0, 1, \dots, 1)$, the ridge objective is

$$F(\theta) = \frac{1}{2N} \|A\theta - y\|_2^2 + \frac{\lambda}{2} \|D\theta\|_2^2.$$

The penalty shrinks slope coefficients but leaves the intercept unpenalized. Its derivatives are

$$\nabla F(\theta) = \frac{1}{N} A^\top (A\theta - y) + \lambda D^\top D\theta,$$

$$\nabla^2 F(\theta) = \frac{1}{N} A^\top A + \lambda D^\top D \succeq 0.$$

Thus the problem is convex. Setting the gradient to zero gives the normal equations

$$\left(\frac{1}{N} A^T A + \lambda D^T D \right) \theta = \frac{1}{N} A^T y.$$

They can be solved directly; a pseudoinverse handles a singular unregularized case.

After keeping functioning days, there are 8,465 chronological rows. The split contains 5,925 training rows, 1,270 validation rows, and 1,270 test rows. Means and standard deviations are computed from the training portion only and then reused for validation and testing. Validation selects $\lambda = 0.01$ from the requested grid. The final test metrics are

$$\text{RMSE} = 457.751, \quad \text{MAE} = 307.486.$$

The training/validation-mean baseline has test RMSE 567.498, so the ridge model improves on the baseline, although the remaining error is substantial.

A chronological split avoids training on observations that occur after the validation or test observations. Randomly shuffling these hourly records could leak seasonal and time-dependent structure. Limitations include nonlinear weather/hour interactions, seasonal drift, unusual events, autocorrelated errors, and unconstrained linear predictions that can become negative even though demand is a count.

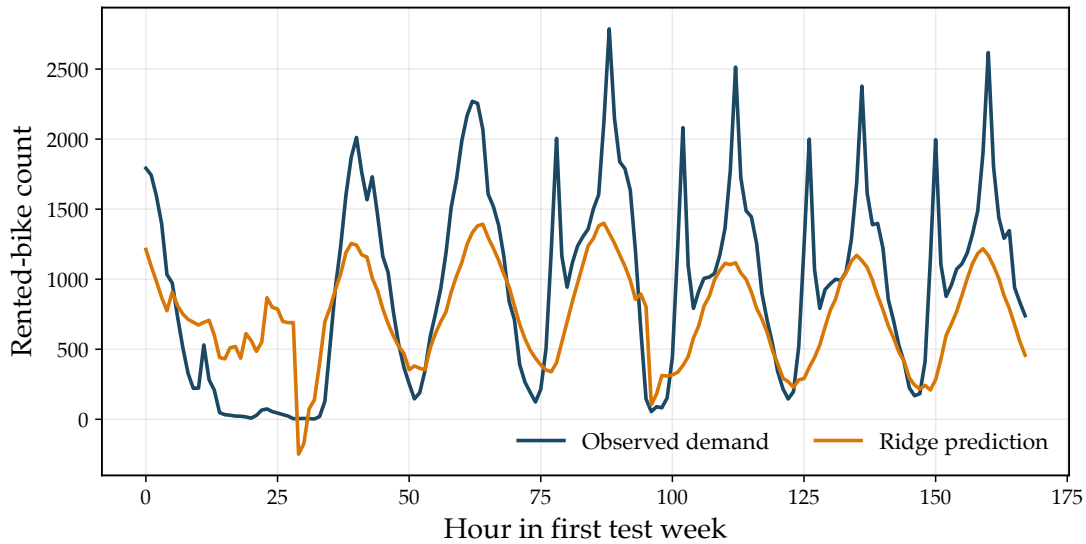


Figure 8: Observed and ridge-predicted demand over the first 168 test hours.

Problem 19: Occupancy detection with logistic regression

Solution. Let $z = A\theta$ and $p_i = \sigma(z_i)$. The regularized average logistic loss is

$$F(\theta) = \frac{1}{N} \sum_i [\log(1 + e^{z_i}) - y_i z_i] + \frac{\lambda}{2} \|D\theta\|_2^2.$$

Since $d \log(1 + e^z) / dz = \sigma(z)$,

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

Writing $W = \text{diag}(p_i(1 - p_i))$ gives

$$\nabla^2 F(\theta) = \frac{1}{N} A^T W A + \lambda D^T D \succeq 0,$$

because $W \succeq 0$. Hence training is a convex optimization problem.

With training-only standardization, $\lambda = 0.01$, and an unpenalized intercept, the fitted coefficient vector is

$$\theta \approx (-2.83395, -0.15248, 0.02101, 2.31525, 0.97966, 0.01531).$$

At threshold 0.5, with rows denoting actual class and columns denoting predicted class, the test confusion matrix is

$$\begin{bmatrix} 1638 & 55 \\ 3 & 969 \end{bmatrix}.$$

Thus

$$\text{accuracy} = \frac{1638 + 969}{2665} = 0.9782, \quad \text{precision} = \frac{969}{969 + 55} = 0.9463, \quad \text{recall} = \frac{969}{969 + 3} = 0.9969.$$

Accuracy alone can be misleading when classes are imbalanced or when the two error types have unequal cost. In HVAC control, a false vacancy can reduce comfort or ventilation for an occupied room, whereas a false occupancy wastes energy. Sensor drift and threshold calibration should also be monitored.

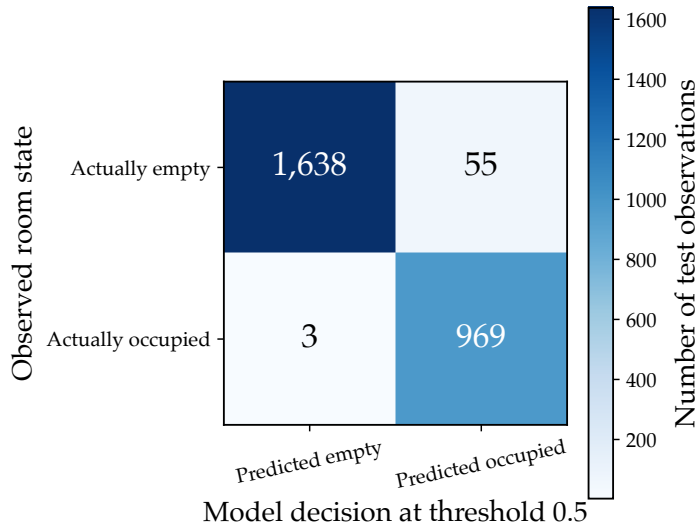


Figure 9: Test confusion matrix for the occupancy classifier.

Reproducibility

Run `code/homework3_fa26_python.py` from the `solutions` directory to reproduce the numerical answers and all figures. The companion file `code/homework3_fa26_matlab.m` provides the MATLAB/CVX implementation.