Lagrange multipliers

Ten problems on optimizing under constraints with Lagrange multipliers: why ∇f = λ∇g, extrema on a line and on a circle, an open box of least area, the closest point on a plane, $x^T Ax$ on the unit sphere, the maximum-entropy and Gibbs distributions, the minimum-norm solution of Ax = b, λ as the sensitivity of the optimum and KKT with active and inactive constraints, with worked solutions and the mistakes that cancel a variable or drop a candidate.

BEFORE YOU START

When a point must stay on a curve or a surface, setting the gradient to zero is replaced by $\nabla f = \lambda \nabla g$: the gradient of the objective lines up with the gradient of the constraint. These ten problems derive that condition, use it on a line, a circle, a box and a plane, then draw from it three results machine learning leans on — eigenvalues as the extremes of a quadratic form, the softmax as a maximum-entropy distribution and the minimum-norm solution of $Ax = b$ — and end with what λ measures and with inequality constraints. The five mistakes each lose a candidate or misread the multiplier.

- f and g are functions of $x \in \mathbb{R}^n$ (in the plane, of (x, y)) with continuous partial derivatives; ∇f is the column of partial derivatives, as on the partial-derivatives page. The constraint is $g(x) = c$. At a point of $g = c$ where f, restricted to the constraint, has a local maximum or minimum and $\nabla g \neq 0$, there is a number λ, the Lagrange multiplier, with $\nabla f = \lambda \nabla g$. The points satisfying this together with $g = c$ are the candidates.
- The same equations come from setting every partial derivative of the Lagrangian $L(x, \lambda) = f(x) - \lambda\big(g(x) - c\big)$ to 0; $\partial L / \partial \lambda = 0$ is the constraint. Writing $f + \lambda(g - c)$ instead negates λ; this page never does.
- With several constraints $g_1 = c_1, \dots, g_m = c_m$ there is one multiplier for each: $\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_m \nabla g_m$.
- Choosing among candidates: on a closed and bounded constraint set, such as a circle or a sphere, f attains a maximum and a minimum, both are candidates, and comparing the values of f picks them out. On an unbounded set, such as a line or a plane, a candidate can be a maximum, a minimum or neither, and needs a separate argument.
- Inequality constraints (Problem 10): to minimize f subject to $h(x) \leq 0$, the Karush–Kuhn–Tucker (KKT) conditions are $\nabla f + \mu \nabla h = 0$, $\mu \geq 0$, $h \leq 0$ and $\mu h = 0$. The last says either $\mu = 0$ (the constraint is inactive, and the point is an ordinary stationary point) or $h = 0$ (it is active, and the point is on the boundary).
- log is the natural log; the entropy of a distribution p on n outcomes is $H(p) = -\sum_i p_i \log p_i$.

Builds on: Partial derivatives, gradients and Hessians

Problems

1. Explain why, at a point x^* of the curve $g(x, y) = c$ where f restricted to the curve has a local maximum or minimum and $\nabla g(x^*) \neq 0$, the gradients satisfy $\nabla f(x^*) = \lambda \nabla g(x^*)$ for some number λ. difficulty ·

2. Find the maximum of $f(x, y) = xy$ on the line $x + 2y = 8$, and the multiplier. Is there a minimum? difficulty ·

3. Find the maximum and minimum of $f(x, y) = x^2 + y$ on the unit circle $x^2 + y^2 = 1$. List every candidate. difficulty ··

4. An open-top box with a rectangular base x by y and height z must hold volume $xyz = 32$. Find the dimensions that minimize the area of base and sides, $S = xy + 2xz + 2yz$. difficulty ··

5. Find the point of the plane $x_1 + 2x_2 + 2x_3 = 14$ closest to $p = (1, 1, 1)^\top$, and its distance from p. difficulty ··

6. For a symmetric matrix A, show that the largest value of $x^\top A x$ over unit vectors x is the largest eigenvalue $\lambda_{\max}(A)$. Find the maximum and minimum for $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. difficulty ··

7. Find the distribution p on n outcomes with the largest entropy when the only constraint is $\sum_i p_i = 1$. Then give the outcomes values $x_1, \dots, x_n$ and also fix the mean, $\sum_i p_i x_i = \mu$: show the maximizer has the form $p_i = e^{-\beta x_i} / \sum_j e^{-\beta x_j}$, and find β and p for $x = (0, 1, 2)$ and $\mu = \frac{4}{7}$. difficulty ···

8. A is $m \times n$ with $m < n$ and independent rows, so $Ax = b$ has infinitely many solutions. Find the one of smallest norm by minimizing $\frac{1}{2}\|x\|^2$ subject to $Ax = b$, and compute it for $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ and $b = (1, 1)^\top$. difficulty ···

9. Replace the constraint of Problem 2 by $x + 2y = c$ with $c > 0$. Find the maximum $f^*(c)$ and the multiplier $\lambda(c)$, and show $\dfrac{df^*}{dc} = \lambda$. Then show the same for any f and g whose optimum $x^*(c)$ moves differentiably with c. difficulty ···

10. Minimize $f(x, y) = (x - 2)^2 + (y - 1)^2$ subject to (a) $x + y \le 2$, and (b) $x + y \le 4$. In each case say whether the constraint is active. difficulty ···

Answers

1. $\nabla f(x^*) = \lambda \nabla g(x^*)$ for some number λ: both gradients are orthogonal to the tangent $r'(0)$, and in the plane those vectors form a single line
2. maximum $f = 8$ at $(4, 2)$, with $\lambda = 2$; there is no minimum
3. maximum $\frac{5}{4}$ at $(\pm\frac{\sqrt{3}}{2}, \frac{1}{2})$; minimum -1 at $(0, -1)$; the candidate $(0, 1)$, with $f = 1$, is neither
4. $x = y = 4$, $z = 2$; least area 48
5. closest point $(2, 3, 3)^\top$; distance 3
6. $\max_{\|x\|=1} x^\top A x = \lambda_{\max}(A)$, attained at a unit eigenvector; here the maximum is 3 at $\pm\frac{1}{\sqrt{2}}(1, 1)^\top$ and the minimum 1 at $\pm\frac{1}{\sqrt{2}}(1, -1)^\top$
7. Normalization only: $p_i = \frac{1}{n}$, $H = \log n$. With the mean: $p_i = \dfrac{e^{-\beta x_i}}{\sum_j e^{-\beta x_j}}$, and for $x = (0, 1, 2)$, $\mu = \frac{4}{7}$: $\beta = \log 2$, $p = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$
8. $x^* = A^\top (A A^\top)^{-1} b$; for this A and b, $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3})^\top$
9. $f^*(c) = \frac{c^2}{8}$ and $\lambda(c) = \frac{c}{4}$, so $\dfrac{df^*}{dc} = \frac{c}{4} = \lambda$; in general $\dfrac{df^*}{dc} = \nabla f^\top \dfrac{dx^*}{dc} = \lambda \, \nabla g^\top \dfrac{dx^*}{dc} = \lambda$
10. (a) active: $(\frac{3}{2}, \frac{1}{2})$, $\mu = 1$, $f = \frac{1}{2}$; (b) inactive: $(2, 1)$, $\mu = 0$, $f = 0$

Worked solutions

Problem 1

> Explain why, at a point x^* of the curve $g(x, y) = c$ where f restricted to the curve has a local maximum or minimum and $\nabla g(x^*) \neq 0$, the gradients satisfy $\nabla f(x^*) = \lambda \nabla g(x^*)$ for some number λ.

Near x^* the curve is traced by a path $r(t)$ with $r(0) = x^*$ and a nonzero velocity $r'(0)$, which is tangent to the curve.	$\nabla g(x^*) \neq 0$ guarantees this: the implicit function theorem then solves $g = c$ for one coordinate in terms of the other near x^*, so the curve is a smooth graph there.
$g(r(t)) = c$ for every small t, so $\nabla g(x^*)^\top r'(0) = 0$.	The path stays on the curve, so g is constant along it; differentiate with the chain rule, $\frac{d}{dt} g(r(t)) = \nabla g(r(t))^\top r'(t)$, at $t = 0$.
$t \mapsto f(r(t))$ has a local maximum or minimum at $t = 0$, so $\nabla f(x^*)^\top r'(0) = 0$.	A differentiable function of one variable has derivative 0 at an interior local extremum; the same chain rule gives the derivative.
In the plane, the vectors orthogonal to the nonzero vector $r'(0)$ form a line through the origin, and $\nabla g(x^*) \neq 0$ spans it.	A line through the origin is spanned by any nonzero vector on it, and step 2 puts $\nabla g(x^*)$ on it.
$\nabla f(x^*) = \lambda \nabla g(x^*)$ **for some number** λ**: both gradients are orthogonal to the tangent** $r'(0)$**, and in the plane those vectors form a single line**	Step 3 puts $\nabla f(x^*)$ on the same line as $\nabla g(x^*)$, so it is a multiple of it. Geometrically, the level curve of f through x^* touches the constraint curve there: if ∇f had a component along the tangent, moving along the curve in that direction would increase f. In $\mathbb{R}^n$ the same argument runs over every tangent direction of the surface $g = c$.

Problem 2

> Find the maximum of $f(x, y) = xy$ on the line $x + 2y = 8$, and the multiplier. Is there a minimum?

$\nabla f = (y, x)^\top$ and $\nabla g = (1, 2)^\top$ for $g(x, y) = x + 2y$.	$\partial(xy)/\partial x = y$ and $\partial(xy)/\partial y = x$; g is linear, so its partial derivatives are its coefficients.
$y = \lambda$ and $x = 2\lambda$.	$\nabla f = \lambda \nabla g$, one equation per component.
$2\lambda + 2 \cdot \lambda = 8$, so $\lambda = 2$ and $(x, y) = (4, 2)$, with $f = 8$.	Substitute step 2 into the constraint $x + 2y = 8$: the third equation, which fixes λ.
Along the line $x = 8 - 2y$, so $f = (8 - 2y)y = 8 - 2(y - 2)^2$.	Parametrize the constraint by y, then complete the square: $8y - 2y^2 = 8 - 2(y^2 - 4y + 4)$.
maximum $f = 8$ at $(4, 2)$, with $\lambda = 2$; there is no minimum	Step 4 is a downward parabola in y: it peaks at $y = 2$ with value 8 and goes to $-\infty$ as $\lvert y \rvert$ grows. The line is unbounded, so the single candidate needed this separate argument. At $(4, 2)$, $\nabla f = (2, 4)^\top = 2\,(1, 2)^\top$.

Problem 3

> Find the maximum and minimum of $f(x, y) = x^2 + y$ on the unit circle $x^2 + y^2 = 1$. List every candidate.

$\nabla f = (2x, 1)^\top$ and $\nabla g = (2x, 2y)^\top$ for $g(x, y) = x^2 + y^2$.	Partial derivatives term by term.
$2x = 2\lambda x$, $1 = 2\lambda y$ and $x^2 + y^2 = 1$.	$\nabla f = \lambda \nabla g$ by components, and the constraint.
$2x(1 - \lambda) = 0$, so $x = 0$ or $\lambda = 1$.	Move everything to one side and factor rather than divide by x, which may be 0. A product is 0 when either factor is, so both cases must be followed.

Case $x = 0$: $y^2 = 1$, so $y = 1$ or $y = -1$, with $\lambda = \frac{1}{2}$ or $-\frac{1}{2}$.	The constraint, then $\lambda = 1/(2y)$ from the second equation. $y^2 = 1$ has two roots and both are candidates.
Case $\lambda = 1$: $1 = 2y$ gives $y = \frac{1}{2}$, and then $x^2 = \frac{3}{4}$, so $x = \pm\frac{\sqrt{3}}{2}$.	The second equation, then the constraint.
$f(0, 1) = 1$, $f(0, -1) = -1$ and $f(\pm\frac{\sqrt{3}}{2}, \frac{1}{2}) = \frac{3}{4} + \frac{1}{2} = \frac{5}{4}$.	Evaluate f at all four candidates.
maximum $\frac{5}{4}$ at $(\pm\frac{\sqrt{3}}{2}, \frac{1}{2})$; minimum -1 at $(0, -1)$; the candidate $(0, 1)$, with $f = 1$, is neither	The circle is closed and bounded, so the maximum and minimum exist and are among the four candidates; comparing values picks them. Near $(0, 1)$ the circle is $y \approx 1 - \frac{x^2}{2}$, so $f \approx 1 + \frac{x^2}{2}$: that candidate is only a local minimum along the circle.

Problem 4

> An open-top box with a rectangular base x by y and height z must hold volume $xyz = 32$. Find the dimensions that minimize the area of base and sides, $S = xy + 2xz + 2yz$.

$\nabla S = (y + 2z,\ x + 2z,\ 2x + 2y)^\top$ and $\nabla g = (yz,\ xz,\ xy)^\top$ for $g = xyz$.	Partial derivatives, holding the other two variables fixed.
$y + 2z = \lambda yz$, $x + 2z = \lambda xz$ and $2x + 2y = \lambda xy$.	$\nabla S = \lambda \nabla g$ by components.
$xy + 2xz = \lambda xyz$, $xy + 2yz = \lambda xyz$ and $2xz + 2yz = \lambda xyz$.	Multiply the three equations by x, y and z: every right side becomes λxyz, and nothing has been divided.
$2xz = 2yz$, so $x = y$.	Subtract the second equation from the first. $xyz = 32$ makes every dimension nonzero, so dividing by $2z$ is safe.

$xy = 2yz$, so $x = 2z$.	Subtract the third equation from the first, then divide by $y \neq 0$.
$xyz = 2z \cdot 2z \cdot z = 4z^3 = 32$, so $z = 2$ and $x = y = 4$.	Substitute steps 4 and 5 into the constraint.
$\lambda = \dfrac{y + 2z}{yz} = \dfrac{8}{8} = 1.$	The first equation of step 2; Problem 9 uses this value.
$x = y = 4$, $z = 2$; **least area** 48	$S = 16 + 16 + 16$. It is the minimum: with $z = 32/(xy)$ the area is $xy + 64/y + 64/x$, which grows without bound as x or y goes to 0 or to ∞, so a minimum exists inside and must be a candidate, and this is the only one. The base is square and the height is half its side.

Problem 5

> Find the point of the plane $x_1 + 2x_2 + 2x_3 = 14$ closest to $p = (1, 1, 1)^\top$, and its distance from p.

Write the plane as $a^\top x = 14$ with $a = (1, 2, 2)^\top$.

Minimize $\|x - p\|^2$ subject to $a^\top x = 14$.	Squaring is increasing on nonnegative numbers, so the square has the same minimizer as the distance and no square root to differentiate.
$\nabla \|x - p\|^2 = 2(x - p)$ and $\nabla (a^\top x) = a$.	$\|x - p\|^2 = \sum_i (x_i - p_i)^2$ differentiates term by term; $a^\top x$ is linear.
$2(x - p) = \lambda a$, so $x = p + \frac{\lambda}{2} a$.	Stationarity: the closest point is reached from p by moving along the normal a of the plane.
$a^\top x = a^\top p + \frac{\lambda}{2} a^\top a = 5 + \frac{9}{2}\lambda = 14$, so $\lambda = 2$.	Substitute into the constraint; $a^\top p = 1 + 2 + 2$ and $a^\top a = 1 + 4 + 4$.

$x^* = (1,1,1)^\top + (1,2,2)^\top = (2,3,3)^\top$, and $\|x^* - p\| = \|a\| = 3$.	$\frac{\lambda}{2} = 1$, so $x^* - p = a$; check $a^\top x^* = 2 + 6 + 6 = 14$.
closest point $(2,3,3)^\top$; **distance** 3	It is the minimum: any other point of the plane is $x^* + z$ with $a^\top z = 0$, and $z \perp a = x^* - p$ gives $\|x^* + z - p\|^2 = 9 + \|z\|^2$. In general step 4 gives distance $\|d - a^\top p\|/\|a\|$ for the plane $a^\top x = d$, here $9/3$.

Problem 6

> For a symmetric matrix A, show that the largest value of $x^\top A x$ over unit vectors x is the largest eigenvalue $\lambda_{\max}(A)$. Find the maximum and minimum for $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

$f(x) = x^\top A x$ and $g(x) = x^\top x$ with $c = 1$; $\nabla f = 2Ax$ and $\nabla g = 2x$.	The gradient of $x^\top A x$ is $(A + A^\top)x$, which is $2Ax$ for symmetric A; $x^\top x$ is the case $A = I$.
$2Ax = \lambda \cdot 2x$, that is $Ax = \lambda x$.	Stationarity: the candidates are the unit eigenvectors of A, and the multiplier is the eigenvalue.
At a candidate, $f = x^\top A x = \lambda\, x^\top x = \lambda$.	Substitute $Ax = \lambda x$, then the constraint $x^\top x = 1$.
The unit sphere is closed and bounded, so the maximum exists and is a candidate; the largest candidate value is $\lambda_{\max}(A)$.	By step 3 the candidate values are exactly the eigenvalues; likewise the minimum is the smallest eigenvalue.
For the given A: $\det(A - \lambda I) = (2 - \lambda)^2 - 1 = 0$, so $\lambda = 3$ or 1, with eigenvectors $(1,1)^\top$ and $(1,-1)^\top$.	$2 - \lambda = \pm 1$; then $A(1,1)^\top = (3,3)^\top$ and $A(1,-1)^\top = (1,-1)^\top$.

$\max_{\|x\|=1} x^\top A x = \lambda_{\max}(A)$, **attained at a unit eigenvector; here the maximum is** 3 **at** $\pm\frac{1}{\sqrt{2}}(1,1)^\top$ **and the minimum** 1 **at** $\pm\frac{1}{\sqrt{2}}(1,-1)^\top$	Check: $x = \frac{1}{\sqrt{2}}(1,1)^\top$ gives $x^\top A x = \frac{1}{2}(2+1+1+2) = 3$. This is why the first principal component, the direction of largest variance $x^\top \Sigma x$, is the top eigenvector of the covariance matrix.

Problem 7

> Find the distribution p on n outcomes with the largest entropy when the only constraint is $\sum_i p_i = 1$. Then give the outcomes values $x_1, \dots, x_n$ and also fix the mean, $\sum_i p_i x_i = \mu$: show the maximizer has the form $p_i = e^{-\beta x_i} / \sum_j e^{-\beta x_j}$, and find β and p for $x = (0, 1, 2)$ and $\mu = \frac{4}{7}$.

Every p_i is taken positive. The derivative of $-t \log t$ tends to $+\infty$ as $t \to 0^+$, so moving a little probability onto an empty outcome always raises H, and the maximum is not on the boundary.

$\dfrac{\partial H}{\partial p_i} = -\log p_i - 1.$	Product rule on $-p_i \log p_i$; no other term of H contains p_i.
Normalization only: $-\log p_i - 1 = \lambda$ for every i.	$\nabla H = \lambda \nabla g$ with $g = \sum_i p_i$, whose gradient is the all-ones vector.
$p_i = e^{-1-\lambda}$, the same for every i, so $p_i = \frac{1}{n}$ and $H = -\sum_i \frac{1}{n} \log \frac{1}{n} = \log n$.	Solve step 2 for p_i; then $\sum_i p_i = 1$ fixes the common value.
This is the maximum.	$-t \log t$ has second derivative $-1/t < 0$, so H is concave, and on a constraint set cut out by linear equations a stationary point of a concave function is its maximum. The entropy page proved the same bound with the KL divergence.
With the mean: $-\log p_i - 1 = \lambda + \beta x_i$.	Two constraints, two multipliers: $\nabla H = \lambda \nabla(\sum_i p_i) + \beta \nabla(\sum_i p_i x_i)$, and the second gradient is the vector of values x_i.
$p_i = e^{-1-\lambda} e^{-\beta x_i}$, and $\sum_i p_i = 1$ gives $e^{-1-\lambda} = 1 / \sum_j e^{-\beta x_j}$.	Exponentiate step 5; the factor $e^{-1-\lambda}$ is the same for every i, so normalization determines it.

For $x = (0, 1, 2)$ write $r = e^{-\beta}$: $p = \dfrac{(1, r, r^2)}{1 + r + r^2}$, with mean $\dfrac{r + 2r^2}{1 + r + r^2} = \dfrac{4}{7}$.	$e^{-\beta x_i} = r^{x_i}$, and the mean is $\sum_i p_i x_i$; the remaining constraint fixes β.
$7r + 14r^2 = 4 + 4r + 4r^2$, so $10r^2 + 3r - 4 = (2r - 1)(5r + 4) = 0$, and $r = \frac{1}{2}$.	Cross-multiply and collect terms; $r = e^{-\beta} > 0$ rules out $-\frac{4}{5}$.
$\beta = \log 2$ and $p = \dfrac{(1, \frac{1}{2}, \frac{1}{4})}{7/4} = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$.	$e^{-\beta} = \frac{1}{2}$; $1 + \frac{1}{2} + \frac{1}{4} = \frac{7}{4}$. Check: mean $\frac{2}{7} + \frac{2}{7} = \frac{4}{7}$.
Normalization only: $p_i = \frac{1}{n}$, $H = \log n$. **With the mean:** $p_i = \dfrac{e^{-\beta x_i}}{\sum_j e^{-\beta x_j}}$, **and for** $x = (0, 1, 2)$, $\mu = \frac{4}{7}$: $\beta = \log 2$, $p = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$	The second form is the softmax of $-\beta x$, the Gibbs distribution; $\beta = 0$ gives back the uniform distribution, whose mean here is 1. A mean below 1 needs $\beta > 0$, which tilts the probability toward small x_i. The concavity argument of step 4 again makes it the maximum.

Problem 8

> A is $m \times n$ with $m < n$ and independent rows, so $Ax = b$ has infinitely many solutions. Find the one of smallest norm by minimizing $\frac{1}{2}\|x\|^2$ subject to $Ax = b$, and compute it for $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ and $b = (1, 1)^\top$.

The constraints are $a_i^\top x = b_i$ for $i = 1, \dots, m$, where $a_i^\top$ is row i of A; give them multipliers $\nu_1, \dots, \nu_m$, collected in $\nu \in \mathbb{R}^m$.	One multiplier per constraint; ν avoids a clash with the eigenvalue λ of Problem 6.
$x = \sum_i \nu_i a_i = A^\top \nu$.	Stationarity: $\nabla \frac{1}{2}\|x\|^2 = x$ and $\nabla(a_i^\top x) = a_i$. The columns of $A^\top$ are the a_i, so the sum is $A^\top \nu$.
$AA^\top \nu = b$, so $\nu = (AA^\top)^{-1} b$.	Substitute into $Ax = b$. $AA^\top$ is $m \times m$ and invertible because the rows of A are independent.

$x^* = A^\top (AA^\top)^{-1} b$.	Substitute ν back into step 2.
Any other solution is $x^* + z$ with $Az = 0$, and $\|x^* + z\|^2 = \|x^*\|^2 + 2\nu^\top Az + \|z\|^2 = \|x^*\|^2 + \|z\|^2$.	${x^*}^\top z = \nu^\top Az = 0$, so the cross term vanishes; every other solution is longer, and x^* is the minimum.
For the given A: $AA^\top = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $(AA^\top)^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, and $\nu = \frac{1}{3}(1,1)^\top$.	Entry (i,j) of $AA^\top$ is row i dotted with row j; the determinant is 3, and the 2×2 inverse swaps the diagonal and negates the off-diagonal.
$x^* = A^\top \nu = \frac{1}{3}(1,0,1)^\top + \frac{1}{3}(0,1,1)^\top = \frac{1}{3}(1,1,2)^\top$.	$A^\top \nu$ is ν_1 times the first row of A plus ν_2 times the second. Check: $Ax^* = \frac{1}{3}(1+2,\ 1+2)^\top = (1,1)^\top$.
$x^* = A^\top (AA^\top)^{-1} b$; **for this A and b,** $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3})^\top$	x^* lies in the row space of A, orthogonal to the null space: here $(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}) \cdot (1,1,-1) = 0$. It is the pseudoinverse solution $A^+ b$ of the least-squares page, the wide-matrix counterpart of $(A^\top A)^{-1} A^\top b$.

Problem 9

> Replace the constraint of Problem 2 by $x + 2y = c$ with $c > 0$. Find the maximum $f^*(c)$ and the multiplier $\lambda(c)$, and show $\dfrac{df^*}{dc} = \lambda$. Then show the same for any f and g whose optimum $x^*(c)$ moves differentiably with c.

$y = \lambda$, $x = 2\lambda$ and $4\lambda = c$, so $\lambda(c) = \frac{c}{4}$ and $(x, y) = (\frac{c}{2}, \frac{c}{4})$.	Problem 2's equations with 8 replaced by c.
$f^*(c) = \frac{c}{2} \cdot \frac{c}{4} = \frac{c^2}{8}$, so $\dfrac{df^*}{dc} = \frac{c}{4} = \lambda(c)$.	Differentiate in c; at $c = 8$ both are 2, as in Problem 2.

In general $f^*(c) = f(x^*(c))$, so $\dfrac{df^*}{dc} = \nabla f(x^*)^\top \dfrac{dx^*}{dc}$.	Chain rule through the path the optimum traces as c changes.
$\nabla f(x^*)^\top \dfrac{dx^*}{dc} = \lambda \nabla g(x^*)^\top \dfrac{dx^*}{dc}$.	Stationarity at the optimum, $\nabla f = \lambda \nabla g$.
$g(x^*(c)) = c$ for every c, so $\nabla g(x^*)^\top \dfrac{dx^*}{dc} = 1$.	Differentiate the constraint, which holds identically in c, with the same chain rule.
$f^*(c) = \frac{c^2}{8}$ **and** $\lambda(c) = \frac{c}{4}$, **so** $\dfrac{df^*}{dc} = \frac{c}{4} = \lambda$; **in general** $\dfrac{df^*}{dc} = \nabla f^\top \dfrac{dx^*}{dc} = \lambda \nabla g^\top \dfrac{dx^*}{dc} = \lambda$	λ is the price of the constraint: raising c from 8 to 9 raises the maximum by about 2 (exactly $\frac{81}{8} - 8 = \frac{17}{8}$). For the box of Problem 4, $\lambda = 1$: the least area for volume V is $S^*(V) = 3(2V)^{2/3}$, whose derivative at $V = 32$ is 1, so one more unit of volume costs about one more unit of area.

Problem 10

> Minimize $f(x, y) = (x-2)^2 + (y-1)^2$ subject to (a) $x + y \le 2$, and (b) $x + y \le 4$. In each case say whether the constraint is active.

$h(x, y) = x + y - c \le 0$ with $c = 2$ or 4; $\nabla f = \big(2(x-2),\ 2(y-1)\big)^\top$ and $\nabla h = (1, 1)^\top$.	The KKT conditions of Before you start need the constraint in the form $h \le 0$.
$2(x-2) + \mu = 0$ and $2(y-1) + \mu = 0$, so $x = 2 - \frac{\mu}{2}$ and $y = 1 - \frac{\mu}{2}$.	Stationarity, $\nabla f + \mu \nabla h = 0$, by components.
Inactive case, $\mu = 0$: $(x, y) = (2, 1)$, where $x + y = 3$.	Complementary slackness allows $\mu = 0$; then the point is the unconstrained minimum of f.

(a) $3 > 2$, so $(2, 1)$ is infeasible and the constraint must be active: $x + y = 3 - \mu = 2$ gives $\mu = 1$, the point $(\frac{3}{2}, \frac{1}{2})$ and $f = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.	With the inactive case ruled out, $\mu h = 0$ forces $h = 0$; and $\mu = 1 \geq 0$, so all four conditions hold.
(b) $3 \leq 4$, so $(2, 1)$ is feasible, $\mu = 0$ satisfies all four conditions, and $f = 0$.	$f \geq 0$ everywhere, so 0 cannot be beaten.
(a) active: $(\frac{3}{2}, \frac{1}{2})$, $\mu = 1$, $f = \frac{1}{2}$; **(b) inactive:** $(2, 1)$, $\mu = 0$, $f = 0$	f is convex and the constraint linear, so a KKT point is the minimum. μ is a sensitivity, as in Problem 9: in (a) the minimum is $f^*(c) = \frac{(3-c)^2}{2}$ for $c < 3$, with $\frac{df^*}{dc} = -1 = -\mu$ at $c = 2$, so loosening the constraint lowers the minimum at rate μ; in (b) loosening it changes nothing, and $\mu = 0$.

Where this goes wrong

1. Taking only the positive root of $y^2 = 1$

$\sqrt{1} = 1$ is the habit, and $y^2 = 1$ gets solved as if it were $y = \sqrt{1}$.

In Problem 3 the case $x = 0$ leaves the constraint $y^2 = 1$	Right so far: this is the branch from $2x(1 - \lambda) = 0$.
"Take the square root."	The shortcut that causes the mistake: the square-root sign names one root, and the equation has two.
$y = 1$, so the candidates are $(\pm\frac{\sqrt{3}}{2}, \frac{1}{2})$ with $f = \frac{5}{4}$ and $(0, 1)$ with $f = 1$, and the minimum is 1 ✗	$y^2 = 1$ also has $y = -1$, and $(0, -1)$ gives $f = -1$, the true minimum. Comparing values finds the maximum and minimum only among the candidates actually listed, so a dropped root can silently replace the answer with a merely local extremum, as $(0, 1)$ is here.

2. Dividing by x and losing the x = 0 candidates

Canceling a common factor from both sides is the first move in solving most equations.

Problem 3's conditions are $2x = 2\lambda x$, $1 = 2\lambda y$ and $x^2 + y^2 = 1$	Right so far: stationarity and the constraint.
"Cancel the common factor $2x$."	The habit that causes the mistake: it is valid only when the factor cannot be 0.
$\lambda = 1$, so $y = \frac{1}{2}$ and the only candidates are $(\pm\frac{\sqrt{3}}{2}, \frac{1}{2})$, both with $f = \frac{5}{4}$ ✗	Dividing by x assumes $x \neq 0$; the equation $2x(1-\lambda) = 0$ also holds when $x = 0$, and that branch holds $(0, -1)$, the minimum -1, and $(0, 1)$. The warning sign was there: the circle is closed and bounded, so a minimum must exist, and this list has nothing below $\frac{5}{4}$. In Problem 4 the same division was safe because $xyz = 32$ rules out zeros.

3. Choosing λ instead of solving the constraint

After stationarity, λ looks like a free parameter, and it is tempting to set it for convenience.

In Problem 5, $2(x - p) = \lambda a$, so $x = p + \frac{\lambda}{2}a$	Right so far: the stationarity condition of Problem 5.
"Pick the λ that makes the objective smallest."	The shortcut that causes the mistake: treating λ as a design choice rather than an unknown.
$\lambda = 0$ gives $x = p = (1, 1, 1)^\top$, at distance 0 ✗	p is not on the plane: $a^\top p = 5 \neq 14$. λ is fixed by the constraint equation $a^\top x = 14$, which stationarity alone does not contain; solving it gives $\lambda = 2$, the point $(2, 3, 3)^\top$ and distance 3. Any answer that does not satisfy $g = c$ has solved an unconstrained problem.

4. Reading df*/dc as λ under the other sign convention

Many texts write the Lagrangian as $f + \lambda(g - c)$, and the sensitivity rule gets carried over without its sign.

For Problem 2, $L = xy + \lambda(x + 2y - 8)$; its derivatives give $y = -\lambda$, $x = -2\lambda$, $\lambda = -2$ and the maximizer $(4, 2)$	Right so far: this convention finds the same point, with λ negated.
"λ is the rate at which the optimum changes with c."	The rule that causes the mistake: Problem 9's result, used without the convention it was proved in.
$\lambda = -2$, so raising c from 8 to 9 lowers the maximum of xy by about 2 ✗	With $L = f + \lambda(g - c)$, stationarity says $\nabla f = -\lambda \nabla g$, so $\frac{df^*}{dc} = -\lambda = 2$: the maximum rises. Check directly: $(4.5, 2.25)$ is on $x + 2y = 9$ and gives $xy = \frac{81}{8} > 8$. The sign of λ is a convention; $\frac{df^*}{dc} = \lambda$ holds for $\nabla f = \lambda \nabla g$.

5. Treating an inactive inequality as active

With equality constraints the answer always lies on the constraint, and that expectation carries over to inequalities.

In Problem 10(b), stationarity gives $2(x - 2) + \mu = 0$ and $2(y - 1) + \mu = 0$	Right so far: the KKT stationarity condition.
"The answer of a constrained problem lies on the constraint."	The analogy that causes the mistake: true for $g = c$, not for $h \le 0$.
Set $x + y = 4$: then $\mu = -1$, the point is $(\frac{5}{2}, \frac{3}{2})$, and the minimum is $\frac{1}{2}$ ✗	$\mu = -1$ breaks $\mu \ge 0$: it says moving into the region lowers f, so the boundary is not where the minimum is. The unconstrained minimum $(2, 1)$ has $x + y = 3 < 4$, so it is feasible, with $\mu = 0$ and $f = 0 < \frac{1}{2}$. Always try the inactive case first, and accept the active one only with $\mu \ge 0$.

More at eulerpad.app. This is one of 82 free practice sets with worked solutions at eulerpad.app/practice, alongside guides to the topics behind them.

EulerPad, the app, reads your handwritten math on iPad, iPhone and Android and checks it line by line. It's free to start.

eulerpad.app/practice/lagrange-multipliers