1 Constrained least-squares formulation

For each problem below, identify the decision variables and formulate the corresponding optimization problem using an unweighted least-squares loss. Do not solve the problem.

Worked example 1.1

Given data (x_i, y_i) for $i = 1, \dots, m$, fit an affine function $y = ax + b$ subject to the constraints $a \geq 0$ and $a + b \leq 2$.

Solution. The decision variables are the slope a and intercept b. The residual at the ith data point is $ax_i + b - y_i$, so the constrained least-squares problem is

$$\min_{a,b} \sum_{i=1}^{m} (ax_i + b - y_i)^2 \qquad \text{s.t.} \qquad a \geq 0, \quad a + b \leq 2.$$

Question 1

Linear regression. Given data (x_i, y_i) for $i = 1, \dots, m$, fit the straight line

$$y = ax + b$$

that best matches the observations.

Solution

The decision variables are $a, b \in \mathbb{R}$. The residual at sample i is $ax_i + b - y_i$, so the problem is

$$\min_{a,b \in \mathbb{R}} \sum_{i=1}^{m} (ax_i + b - y_i)^2.$$

Question 2

Polynomial regression. Given data (x_i, y_i) for $i = 1, \dots, m$, fit a polynomial of degree at most d,

$$p(x) = c_0 + c_1 x + \cdots + c_d x^d,$$

that best matches the observations. (Hint: This problem is actually still *linear* regression.)

Solution

The decision variables are the coefficients $c = (c_0, \dots, c_d) \in \mathbb{R}^{d+1}$. The least-squares problem is

$$\min_{c_0, \dots, c_d} \sum_{i=1}^{m} \left(\sum_{j=0}^{d} c_j x_i^j - y_i \right)^2.$$

Although the polynomial is nonlinear in x, it is affine in the decision variables c_j.

Question 3

Constrained nonlinear least-squares. Given data (x_i, y_i) for $i = 1, \dots, m$, fit the nonlinear model

$$y = ae^{-bx} + c$$

subject to

$$a \geq 0, \qquad b \geq 0, \qquad a + c \leq 10.$$

Solution

The decision variables are $a, b, c \in \mathbb{R}$. The constrained nonlinear least-squares problem is

$$\min_{a,b,c} \sum_{i=1}^{m} \left(ae^{-bx_i} + c - y_i\right)^2 \qquad \text{s.t.} \qquad a \geq 0, \quad b \geq 0, \quad a + c \leq 10.$$

Question 4

Nonnegative matrix factorization. A data matrix $M \in \mathbb{R}^{n\times n}$ is believed to be approximately representable as

$$M \approx UV^T, \qquad U_{i,j} \geq 0, \quad V_{i,j} \geq 0 \text{ for all } i, j.$$

Formulate the least-squares problem for finding the factors U and V.

Solution

For a chosen factor width r, take $U, V \in \mathbb{R}^{n\times r}$ as the decision variables. The formulation is

$$\min_{U,V} \; \|UV^T - M\|_F^2 \qquad \text{s.t.} \qquad U_{ij} \geq 0, \quad V_{ij} \geq 0 \quad \text{for all } i, j.$$

Question 5

Low-rank matrix completion. An unknown matrix $M \in \mathbb{R}^{n\times n}$ is known to have rank at most r. Only the entries M_{ij} for $(i, j) \in \Omega$ have been observed. Estimate the missing entries by fitting a rank-r matrix model $X = LR^T$ whose observed entries best match the data.

Solution

Use $L, R \in \mathbb{R}^{n\times r}$ as the decision variables. Since LR^T has rank at most r, the required problem is

$$\min_{L,R\in\mathbb{R}^{n\times r}} \sum_{(i,j)\in\Omega} \left((LR^T)_{ij} - M_{ij}\right)^2.$$

The fitted matrix is $X = LR^T$.

2 Modern optimizers

Many machine-learning objectives have the finite-sum form $f(x) = \frac{1}{m}\sum_{i=1}^{m} f_i(x)$. Gradient descent uses $x^+ = x - \eta\nabla f(x)$. One epoch of SGD processes $f_1, \ldots, f_m$ in the listed order and applies $x \leftarrow x - \eta\nabla f_i(x)$ after each sample. For SGD with momentum, use

$$v \leftarrow \beta v + (1-\beta)\nabla f_i(x), \qquad x \leftarrow x - \eta v.$$

Carry out the requested update exactly. The arithmetic is intended to be done by hand.

Worked example 2.1

Let

$$f_1(x) = f_2(x) = \tfrac{1}{2}(x-2)^2.$$

Starting from $x_0 = 0$ and $v_0 = 0$, perform one epoch of SGD with momentum using $\eta = 1$ and $\beta = \frac{1}{2}$.

Solution. For the first sample, $g_1 = x_0 - 2 = -2$, so $v_1 = \frac{1}{2}(-2) = -1$ and $x_1 = 0 - (-1) = 1$. For the second sample, $g_2 = x_1 - 2 = -1$, so $v_2 = \frac{1}{2}(-1) + \frac{1}{2}(-1) = -1$ and $x_2 = 1 - (-1) = 2$.

Let

$$f_1(x) = \tfrac{1}{2}(x-4)^2, \quad f_2(x) = \tfrac{1}{2}(x-3)^2,$$

Starting from $x_0 = 0$ and $v_0 = 0$:

Question 1

Perform one gradient-descent update with $\eta = \frac{1}{2}$.

Solution

The averaged objective has gradient

$$\nabla f(x) = \tfrac{1}{2}\big[(x-4) + (x-3)\big] = x - \tfrac{7}{2}.$$

Thus $\nabla f(0) = -7/2$ and

$$x_1 = 0 - \tfrac{1}{2}(-\tfrac{7}{2}) = \tfrac{7}{4}.$$

Question 2

Perform one epoch of SGD with $\eta = \frac{1}{2}$ in the order $1, 2$.

Solution

For sample 1, $g_1 = x_0 - 4 = -4$, so $x_1 = 0 - \frac{1}{2}(-4) = 2$. For sample 2, $g_2 = x_1 - 3 = -1$, so

$$x_2 = 2 - \tfrac{1}{2}(-1) = \tfrac{5}{2}.$$

Question 3

Perform one epoch of SGD with momentum using $\eta = \frac{1}{2}$ and $\beta = \frac{1}{2}$ in the order $1, 2$.

Solution

Sample 1 gives $(g_1, v_1, x_1) = (-4, -2, 1)$. Sample 2 then gives $(g_2, v_2, x_2) = (-2, -2, 2)$.

3 Newton's method

For each instance, use the method and parameter μ specified. For a log-barrier instance, add $-\mu \log g(x)$ for every inequality written as $g(x) > 0$. For a quadratic-penalty instance with equality residual $h(x) = 0$, use $\frac{1}{2\mu}\|h(x)\|_2^2$. Barrier starting points are strictly feasible; penalty starting points need not satisfy the equality constraints. For every instance:

(a) Write the unconstrained objective F_μ.

(b) Derive its gradient and Hessian, then evaluate both at the specified starting point.

(c) Perform one full Newton iteration from the specified starting point.

Worked example 3.1

Use the log barrier with $\mu = \frac{1}{2}$ for

$$\min\ \tfrac{1}{2}(x_1 - 2)^2 + \tfrac{1}{2}(x_2 - 1)^2 \qquad \text{s.t.} \qquad x_1 \geq 0,$$

starting from $x_0 = (1, 1)$.

(a) The unconstrained objective is $F_\mu(x) = \frac{1}{2}(x_1 - 2)^2 + \frac{1}{2}(x_2 - 1)^2 - \frac{1}{2}\log x_1$.

(b) Its derivatives are

$$\nabla F_\mu(x) = \begin{bmatrix} x_1 - 2 - (2x_1)^{-1} \\ x_2 - 1 \end{bmatrix}, \quad \nabla^2 F_\mu(x) = \begin{bmatrix} 1 + (2x_1^2)^{-1} & 0 \\ 0 & 1 \end{bmatrix}.$$

(c) At $x_0 = (1, 1)$, the Newton step is

$$p = -\nabla^2 F_\mu(x_0)^{-1} \nabla F_\mu(x_0) = -\begin{bmatrix} 3/2 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -3/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

So $x^+ = x_0 + p = (2, 1)$.

Question 1

Log barrier, $\mu = \frac{1}{4}$. min $\frac{1}{2}(x - 2)^2$ subject to $x \geq 0$, starting from $x = 1$.

Solution

The barrier objective and its derivatives are

$$F_\mu(x) = \tfrac{1}{2}(x - 2)^2 - \tfrac{1}{4}\log x, \quad F_\mu'(x) = x - 2 - \frac{1}{4x}, \quad F_\mu''(x) = 1 + \frac{1}{4x^2}.$$

At $x_0 = 1$, $F_\mu'(1) = -5/4$ and $F_\mu''(1) = 5/4$. Hence

$$p = -\frac{F_\mu'(1)}{F_\mu''(1)} = 1, \qquad x^+ = x_0 + p = 2.$$

Question 2

Quadratic penalty, $\mu = \frac{1}{2}$. min $\frac{1}{2}(x - 1)^2$ subject to $x^2 = 1$, starting from 0.

Solution

Starting from $x_0 = 0$, the penalty coefficient is $1/(2\mu) = 1$. Thus

$$F_\mu(x) = \tfrac{1}{2}(x - 1)^2 + (x^2 - 1)^2, \quad F_\mu'(x) = 4x^3 - 3x - 1, \quad F_\mu''(x) = 12x^2 - 3.$$

At $x_0 = 0$, the gradient is -1 and the Hessian is -3. Therefore

$$p = -\frac{-1}{-3} = -\tfrac{1}{3}, \qquad x^+ = -\tfrac{1}{3}.$$

Question 3

(G) Log barrier, $\mu = \frac{1}{4}$. min $(x_1 - 2)^2 + (x_2 - 1)^2$ subject to $x_1 + x_2 \geq 1$ and $x_1 - x_2 + 1 \geq 0$, starting from $(1, 1)$.

Solution

Let $s = x_1 + x_2 - 1$ and $t = x_1 - x_2 + 1$. The barrier objective is

$$F_\mu(x) = (x_1 - 2)^2 + (x_2 - 1)^2 - \tfrac{1}{4}\log s - \tfrac{1}{4}\log t.$$

With $a = (1, 1)^T$ and $b = (1, -1)^T$,

$$\nabla F_\mu(x) = 2\begin{bmatrix} x_1 - 2 \\ x_2 - 1 \end{bmatrix} - \frac{a}{4s} - \frac{b}{4t}, \quad \nabla^2 F_\mu(x) = 2I + \frac{aa^T}{4s^2} + \frac{bb^T}{4t^2}.$$

At $x_0 = (1, 1)$, $s = t = 1$ and

$$\nabla F_\mu(x_0) = \begin{bmatrix} -5/2 \\ 0 \end{bmatrix}, \qquad \nabla^2 F_\mu(x_0) = \tfrac{5}{2}I.$$

Consequently $p = -\nabla^2 F_\mu(x_0)^{-1} \nabla F_\mu(x_0) = (1, 0)$ and $x^+ = (2, 1)$.

Question 4

(G) Quadratic penalty, $\mu = \frac{1}{5}$. min $\frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}x_2^2$ subject to $x_1 + x_2 = 1$, starting from $(0, 0)$.

Solution

Let $h(x) = x_1 + x_2 - 1$. Since $1/(2\mu) = 5/2$,

$$F_\mu(x) = \tfrac{1}{2}(x_1 - 1)^2 + \tfrac{1}{2}x_2^2 + \tfrac{5}{2}h(x)^2.$$

Its derivatives are

$$\nabla F_\mu(x) = \begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix} + 5h(x)\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \nabla^2 F_\mu(x) = \begin{bmatrix} 6 & 5 \\ 5 & 6 \end{bmatrix}.$$

At $x_0 = (0, 0)$, the gradient is $(-6, -5)^T$. Therefore

$$p = -\begin{bmatrix} 6 & 5 \\ 5 & 6 \end{bmatrix}^{-1} \begin{bmatrix} -6 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad x^+ = (1, 0).$$

4 Convex functions

Prove or disprove whether each of the following functions is convex on the stated domain.

How to write your solutions. State a criterion for convexity from lecture, apply it explicitly to the function, and conclude whether the criterion holds. To disprove convexity, it is enough to give one point or one choice of x, y, t for which the required condition fails.

Worked example 4.1

Prove or disprove whether $f(x) = x^2$ is convex.

Solution by Hessian identity. A twice differentiable function is convex if and only if its Hessian is positive semidefinite everywhere on its domain. Here, $f''(x) = 2 > 0$ for every $x \in \mathbb{R}$. Therefore f is convex.

Solution using the definition. A function f is convex if and only if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{*}$$

for all x, y and $0 \leq t \leq 1$. For $f(x) = x^2$, we find that

$$\begin{aligned}
&tf(x) + (1-t)f(y) - f(tx + (1-t)y) \\
=&tx^2 + (1-t)y^2 - [tx^2 + 2t(1-t)xy + (1-t)^2y^2] \\
=&tx^2 + [(1-t) - (1-t)^2]y^2 - 2t(1-t)xy \\
=&t(1-t)x^2 + t(1-t)y^2 - 2t(1-t)xy \\
=&t(1-t)(x-y)^2 \geq 0.
\end{aligned}$$

Therefore (*) holds, so f is convex.

Prove or disprove whether $f(x) = x^3$ is convex.

Solution by Hessian identity. A twice differentiable function is convex if and only if its Hessian is positive semidefinite everywhere on its domain. Here, $f''(x) = 6x$, and $f''(-1) = -6 < 0$. The Hessian of f is not positive semidefinite everywhere, so it is *not* convex.

Question 1

$f(x) = \exp(-x)$ for $x \in \mathbb{R}$.

Solution

Since $f''(x) = e^{-x} > 0$ for every $x \in \mathbb{R}$, the Hessian criterion shows that f is strictly convex, hence convex.

Question 2

$f(x, y) = xy$ for $(x, y) \in \mathbb{R}^2$.

Solution

The Hessian is

$$\nabla^2 f(x, y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Its eigenvalues are 1 and -1, so it is indefinite everywhere. Therefore f is not convex.

Question 3

$f(x, y) = \exp(x)\exp(y)$ for $(x, y) \in \mathbb{R}^2$.

Solution

Writing $f(x, y) = e^{x+y}$, we obtain

$$\nabla^2 f(x, y) = e^{x+y}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

For every $z \in \mathbb{R}^2$, $z^T \nabla^2 f(x, y) z = e^{x+y}(z_1 + z_2)^2 \geq 0$. Hence f is convex.

Question 4

$f(x_1, x_2) = (x_1 - 2x_2)^2 + x_2^2$ for $(x_1, x_2) \in \mathbb{R}^2$.

Solution

The Hessian is

$$\nabla^2 f = \begin{bmatrix} 2 & -4 \\ -4 & 10 \end{bmatrix}.$$

Its leading principal minors are $2 > 0$ and $20 - 16 = 4 > 0$, so the Hessian is positive definite. Thus f is strictly convex.

Question 5

(G) $f(x) = x^4$ for $x \in \mathbb{R}$.

Solution

We have $f''(x) = 12x^2 \geq 0$ for all x. Therefore f is convex on $\mathbb{R}$.

Question 6

(G) $f(x) = \log(\exp(x_1) + \exp(x_2)) + x_3$ for $x \in \mathbb{R}^3$.

Solution

Let $p = e^{x_1}/(e^{x_1} + e^{x_2})$. The Hessian is

$$\nabla^2 f(x) = p(1 - p)\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Its quadratic form is $p(1 - p)(z_1 - z_2)^2 \geq 0$. Hence the log-sum-exp term is convex, and adding the affine term x_3 preserves convexity.

Question 7

(G) $f(x, y) = x^2/y$ for $(x, y) \in \mathbb{R} \times \mathbb{R}_{++}$.

Solution

For $y > 0$,

$$\nabla^2 f(x,y) = \begin{bmatrix} 2/y & -2x/y^2 \\ -2x/y^2 & 2x^2/y^3 \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0.$$

Therefore f is convex on $\mathbb{R} \times \mathbb{R}_{++}$.

Question 8

(G) $f(x,y) = -\log(x^2 - y^T y)$ for $x \in \mathbb{R}$, $y \in \mathbb{R}^n$, on the domain $x > \|y\|_2$.

Solution

Let $s = x^2 - \|y\|_2^2 > 0$. The Hessian is

$$\nabla^2 f(x,y) = \frac{2}{s^2}\begin{bmatrix} x^2 + \|y\|_2^2 & -2xy^T \\ -2xy & sI + 2yy^T \end{bmatrix}.$$

For a direction $(a,b) \in \mathbb{R} \times \mathbb{R}^n$, its quadratic form can be written as

$$\frac{4(xa - y^T b)^2}{s^2} - \frac{2(a^2 - \|b\|_2^2)}{s}.$$

If $a^2 \leq \|b\|_2^2$, this is nonnegative immediately. If $a^2 > \|b\|_2^2$, the reverse Cauchy--Schwarz inequality on the second-order cone gives

$$(xa - y^T b)^2 \geq (x^2 - \|y\|_2^2)(a^2 - \|b\|_2^2) = s(a^2 - \|b\|_2^2),$$

which again makes the quadratic form nonnegative. Thus the Hessian is positive semidefinite throughout the domain, so f is convex.

5 Convex sets

Prove or disprove whether each of the following sets is convex.

How to write your solutions. State a criterion for convexity from lecture, apply it explicitly to the function, and conclude whether the criterion holds. When disproving convexity, give two points in the set whose line segment is not entirely contained in the set.

Worked example 5.1

Prove or disprove whether $C = \{x \in \mathbb{R}^2 : x_1 + x_2 \leq 1\}$ is convex.

Solution by definition. Let $u, v \in C$ and $0 \leq t \leq 1$. Since $u_1 + u_2 \leq 1$ and $v_1 + v_2 \leq 1$,

$$[tu + (1-t)v]_1 + [tu + (1-t)v]_2 = t(u_1 + u_2) + (1-t)(v_1 + v_2) \leq 1.$$

Therefore $tu + (1-t)v \in C$, so C is convex.

Solution by level set. Observe that $C = \{x \in \mathbb{R}^2 : f(x) \leq 1\}$ where $f(x) = x_1 + x_2$. Indeed, f is convex, because

$$\begin{aligned} f(tx + (1-t)y) &= [tx_1 + (1-t)y_1] + [tx_2 + (1-t)y_2] \\ &= t(x_1 + x_2) + (1-t)(y_1 + y_2) = tf(x) + (1-t)f(y). \end{aligned}$$

Therefore, C is convex.

Question 1

$C = [-2, 3] \subset \mathbb{R}$. The is the closed interval from -2 to $+2$.

Solution

The set is convex. If $u, v \in [-2, 3]$ and $0 \leq t \leq 1$, then

$$-2 = t(-2) + (1-t)(-2) \leq tu + (1-t)v \leq t(3) + (1-t)(3) = 3.$$

Thus the entire line segment between u and v remains in C.

Question 2

$C = \{-1, +1\} \subset \mathbb{R}$. This is the set with two elements, -1 and $+1$.

Solution

The set is not convex. Both -1 and 1 belong to C, but their midpoint is 0, which does not belong to C.

Question 3

$C = \{x \in \mathbb{R}^2 : x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + x_2^2 \leq 1\}$. Hint: You may use the fact that x^2 is convex, which we proved as a worked example.

Solution

The set is convex. The constraints $x_1 \geq 0$ and $x_2 \geq 0$ define halfspaces. The function $g(x) = x_1^2 + x_2^2$ is convex because $\nabla^2 g = 2I \succeq 0$, so $\{x : g(x) \leq 1\}$ is a convex sublevel set. Their intersection is convex.

Question 4

(G) $C = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : \|x\|_2 \le t\}$. Hint: Use the triangle inequality $\|x + y\| \le \|x\| + \|y\|$.

Solution

Let $(t, x), (s, y) \in C$ and $0 \le \theta \le 1$. By the triangle inequality,

$$\|\theta x + (1 - \theta)y\|_2 \le \theta\|x\|_2 + (1 - \theta)\|y\|_2 \le \theta t + (1 - \theta)s.$$

Therefore $(\theta t + (1 - \theta)s, \theta x + (1 - \theta)y) \in C$, so C is convex.

Question 5

(G) $C = \{X \in \mathbb{S}^n : X \succeq 0\}$, where $\mathbb{S}^n$ denotes the symmetric $n \times n$ matrices. Hint: Use the fact that $X \succeq 0$ if and only if $a^T X a \ge 0$ for all $a \in \mathbb{R}^n$.

Solution

Let $X, Y \succeq 0$ and $0 \le t \le 1$. For every $a \in \mathbb{R}^n$,

$$a^T(tX + (1 - t)Y)a = t\,a^T X a + (1 - t)a^T Y a \ge 0.$$

Hence $tX + (1 - t)Y \succeq 0$, and the positive-semidefinite cone is convex.

Question 6

(G) $C = \{X \in \mathbb{S}^n : X \succeq 0,\ \operatorname{rank}(X) \le 1\}$ for $n \ge 2$.

Solution

The set is not convex. Let e_1, e_2 be two distinct coordinate vectors and define $X = e_1 e_1^T$ and $Y = e_2 e_2^T$. Both matrices are positive semidefinite and have rank one. However, $(X + Y)/2$ has two positive diagonal entries and therefore rank two. Thus the midpoint does not belong to C.