ryz@illinois.edu

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. Given data (x_i, y_i) for $i = 1, \ldots, m$, 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

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

1. **Linear regression.** Given data (x_i, y_i) for $i = 1, \ldots, m$, fit the straight line
$$y = ax + b$$
that best matches the observations.
2. **Polynomial regression.** Given data (x_i, y_i) for $i = 1, \ldots, m$, 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.)
3. **Constrained nonlinear least-squares.** Given data (x_i, y_i) for $i = 1, \ldots, m$, fit the nonlinear model
$$y = ae^{-bx} + c$$
subject to
$$a \geq 0, \qquad b \geq 0, \qquad a + c \leq 10.$$
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.
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.

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. 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 $v_0 = 0$:

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

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. 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

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

(c) At $x_0 = (1,1)$, we have the Newto step is

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

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

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

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 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

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

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

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

Therefore (*) holds, so f is convex.

Worked example 2. 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.

1. $f(x) = \exp(-x)$ for $x \in \mathbb{R}$.
2. $f(x, y) = xy$ for $(x, y) \in \mathbb{R}^2$.
3. $f(x, y) = \exp(x)\exp(y)$ for $(x, y) \in \mathbb{R}^2$.
4. $f(x_1, x_2) = (x_1 - 2x_2)^2 + x_2^2$ for $(x_1, x_2) \in \mathbb{R}^2$.
5. **(G)** $f(x) = x^4$ for $x \in \mathbb{R}$.
6. **(G)** $f(x) = \log(\exp(x_1) + \exp(x_2)) + x_3$ for $x \in \mathbb{R}^3$.
7. **(G)** $f(x, y) = x^2/y$ for $(x, y) \in \mathbb{R} \times \mathbb{R}_{++}$.
8. **(G)** $f(x, y) = -\log(x^2 - y^T y)$ for $x \in \mathbb{R}$, $y \in \mathbb{R}^n$, on the domain $x > \|y\|_2$.

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

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

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

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

Solution by level set. Observe that $C = \{x \in \mathbb{R}^2 : f(x) \le 1\}$ 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.

1. $C = [-2, 3] \subset \mathbb{R}$. The is the closed interval from -2 to $+2$.
2. $C = \{-1, +1\} \subset \mathbb{R}$. This is the set with two elements, -1 and $+1$.
3. $C = \{x \in \mathbb{R}^2 : x_1 \ge 0,\ x_2 \ge 0,\ x_1^2 + x_2^2 \le 1\}$. Hint: You may use the fact that x^2 is convex, which we proved as a worked example.
4. **(G)** $C = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : \|x\|_2 \le t\}$. Hint: Use the triangle inequality $\|x + y\| \le \|x\| + \|y\|$.
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 $a \in \mathbb{R}^n$.
6. **(G)** $C = \{X \in \mathbb{S}^n : X \succeq 0,\ \mathrm{rank}(X) \le 1\}$ for $n \ge 2$.