Chapter 1

Review of basic definitions and notation

1.1 Vectors and matrices

Vectors. A vector $v \in \mathbb{R}^n$ is a column consisting of n components:

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix},$$

sometimes written $v = (v_1, \ldots, v_n)$. This is to be distinguished from the row $v^T = [v_1 \; \ldots \; v_n]$.

Inner Products. The inner or dot product of $u, v \in \mathbb{R}^n$ is

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i.$$

We also write $u^T v$ for the inner product.

If θ is the angle between u and v then

$$\langle u, v \rangle \;=\; \cos(\theta) \|u\| \|v\|.$$

This yields the *Cauchy-Schwarz* inequality: $\langle u, v \rangle \leq \|u\| \|v\|$.

Outer Products. The outer product of $u, v \in \mathbb{R}^n$ is uv^T, which is an $n \times n$ matrix. To see this, note that uv^T is the multiplication of a column vector u, which is an $n \times 1$ matrix, with a row vector v^T, which is a $1 \times n$ matrix.

Euclidean Length or Norm. For $u \in \mathbb{R}^n$,

$$\|u\| = \sqrt{\langle u, u \rangle} = \left(\sum_i u_i^2 \right)^{1/2}.$$

Sequences. Consider a sequence of vectors $\{x^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$, that is, $x^1, x^2, \ldots \in \mathbb{R}^n$. We say $\{x^k\}$ is *bounded* if for some constant $M > 0$, and each $k = 1, 2, \ldots$,

$$\|x^k\| \leq M.$$

If $\{x^k\}$ is bounded then it must have at least one convergent subsequence with a *cluster* or *limit* point x^*.

Matrices. The vector space of $m \times n$ matrices is written $\mathbb{R}^{m \times n}$. Let $B \in \mathbb{R}^{n \times n}$, then B_{ij} denotes the component of B in row i and column j.

B is *symmetric* if $B = B^T$ (transpose of B).

B is *positive definite* if $\langle u, Bu \rangle > 0$ for each nonzero vector $u \in \mathbb{R}^n$.

B is *positive semi-definite* if $\langle u, Bu \rangle \geq 0$ for each nonzero $u \in \mathbb{R}^n$.

Note that if B is positive definite, then it is also positive semidefinite, but that B may be positive semidefinite without being positive definite.

Example 1.1.1 Consider the matrix

$$B = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 4 \end{bmatrix}.$$

We will prove that B is positive definite. Now

$$u^T B u = (u_1, u_2) \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 4 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u_1^2 + u_1 u_2 + 4u_2^2 = \left(u_1 + \frac{1}{2} u_2 \right)^2 + \frac{15}{4} u_2^2.$$

Obviously $(u_1 + \frac{1}{2}u_2)^2 \geq 0$ and $\frac{15}{4}u_2^2 \geq 0$, so it must be that $u^T B u \geq 0$ for all u. Thus B is positive semidefinite. Now suppose B is *not* positive definite. Then there must exist $u \neq 0$ such that $u^T B u = 0$. Thus $u_1 + \frac{1}{2}u_2 = 0$ and $u_2 = 0$. But $u_2 = 0$ implies that $u_1 + \frac{1}{2}u_2 = u_1 + \frac{1}{2}0 = u_1 = 0$, and so $u = 0$, which is a contradiction. We have shown $u^T B u > 0$ for all $u \neq 0$, thus B is positive definite.

Example 1.1.2 Consider the matrix

$$B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

We will prove that B is positive semidefinite, but is not positive definite. Now

$$u^T B u = (u_1, u_2) \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u_1^2 + 4u_1u_2 + 4u_2^2 = (u_1 + 2u_2)^2 \geq 0.$$

so B is certainly positive semidefinite. B cannot be positive definite because $u = (2, -1)$ satisfies $u^T B u = (u_1 + 2u_2)^2 = (2 + 2(-1))^2 = 0$. So B must be positive semidefinite, but not positive definite.

Example 1.1.3 Consider the matrix

$$B = \begin{bmatrix} -1 & 2 \\ 2 & 2 \end{bmatrix}.$$

We will show that B is not positive semidefinite, and so is not positive definite either. Now

$$u^T B u = (u_1, u_2) \begin{bmatrix} -1 & 2 \\ 2 & 2 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = -u_1^2 + 4u_1u_2 + 2u_2^2.$$

Now if we take $u = (1, 0)$ we get $u^T B u = -1 < 0$ so B is not positive semidefinite.

If B is symmetric then it is positive (semi-)definite if and only if each of its eigenvalues is positive (respectively nonnegative). This fact gives us an easier way of checking the positive definiteness or semidefiniteness of symmetric matrices. A positive definite symmetric matrix has all eigenvalues strictly positive and a positive semidefinite symmetric matrix has all eigenvalues non-negative.

For example, for B from Example 1.1.1 we have λ an eigenvalue if and only if

$$\begin{aligned} & (1-\lambda)(4-\lambda) - \tfrac{1}{2}\left(\tfrac{1}{2}\right) &= 0 \\ \iff \quad & 4 - 5\lambda + \lambda^2 - \tfrac{1}{4} &= 0 \\ \iff \quad & \lambda^2 - 5\lambda + \tfrac{15}{4} &= 0. \end{aligned}$$

Thus, $\lambda = (5 + \sqrt{10})/2$ or $(5 - \sqrt{10})/2$. Now $25 > 10 \Rightarrow 5 > \sqrt{10}$ so both eigenvalues are strictly positive, confirming B is positive definite.

For B from Example 1.1.2 we have λ an eigenvalue if and only if

$$\begin{aligned} & (1-\lambda)(4-\lambda) - 2(2) &= 0 \\ \iff & 4 - 5\lambda + \lambda^2 - 4 &= 0 \\ \iff & \lambda(\lambda - 5) &= 0. \end{aligned}$$

Thus, $\lambda = 0$ or 5. So all eigenvalues are non-negative, (but not strictly positive), confirming B is positive semidefinite (but not positive definite).

For B from Example 1.1.3 we have λ an eigenvalue if and only if

$$\begin{aligned} & (-1-\lambda)(2-\lambda) - 2(2) &= 0 \\ \iff & -2 - \lambda + \lambda^2 - 4 &= 0 \\ \iff & \lambda^2 - \lambda - 6 &= 0 \\ \iff & (\lambda - 3)(\lambda + 2) &= 0 \end{aligned}$$

Thus, $\lambda = -2$ or 3. So not all eigenvalues are non-negative, confirming B is neither positive semidefinite nor positive definite.

If B is a diagonal matrix then its eigenvalues are its diagonal entries, so it is positive (semi-)definite if and only if its diagonal elements are positive (nonnegative resp.).

If B is positive definite then it is invertible.

If $n = 2$ and

$$B = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

then B is positive definite if and only if $a > 0$ and $ac - b^2 > 0$. For 2×2 matrices, this gives a very easy way to check positive definiteness. Observe that all matrices B in Examples 1.1.1, 1.1.2 and 1.1.3 are symmetric. Now for B in Example 1.1.1, we have $a = 1$, $b = \frac{1}{2}$ and $c = 4$, so $a = 1 > 0$ and $ac - b^2 = 4 - \frac{1}{4} = \frac{15}{4} > 0$, so by this fact, B must be positive definite, as we already proved, directly. For B in Example 1.1.2, we have $a = 1$, $b = 2$ and $c = 4$, so $ac - b^2 = 4 - 4 = 0 \not> 0$, so by this fact, B cannot be positive definite, as we already proved, directly. Finally, for B in Example 1.1.3, we have $a = -1$, $b = 2$ and $c = 2$, so $a = -1 \not> 0$, so by this fact, B cannot be positive definite, as we already proved, directly.

If $n = 3$ and

$$B = \begin{bmatrix} a & b & 0 \\ b & c & 0 \\ 0 & 0 & d \end{bmatrix}$$

then it is positive definite if and only if $a > 0$, $ac - b^2 > 0$ and $d > 0$.

Multiplication Reminder. For any vector $x \in \mathbb{R}^m$, vector $y \in \mathbb{R}^n$ and matrix $B \in \mathbb{R}^{m \times n}$, where $n, m \geq 1$, the product $x^T B y \in \mathbb{R}$, that is, is a scalar, so $(x^T B y)^T = x^T B y$, and thus

$$x^T B y = y^T B^T x.$$

1.2 Functions and differentiability

Let f be a real function of n variables, $f : \mathbb{R}^n \to \mathbb{R}$.

Little-oh notation. A function of one variable, $\Delta(t)$, is called $o(t)$ or "little-oh of t" if $\Delta(t)/t \to 0$ as $t \to 0$, $t \neq 0$.

Note that any polynomial function of t with degree greater than 1 will be $o(t)$. For example, take $\Delta(t) = t^3$. Then $\frac{\Delta(t)}{t} = \frac{t^3}{t} = t^2 \to 0$ as $t \to 0$. Any function of t with degree less than or equal to 1 will *not* be $o(t)$, for example, take $\Delta(t) = t^{0.5} = \sqrt{t}$. Then $\frac{\Delta(t)}{t} = \frac{\sqrt{t}}{t} = \frac{1}{\sqrt{t}} \not\to 0$ as $t \to 0$.

Gradients. We say f is *differentiable* if for each $x \in \mathbb{R}^n$, there is a vector, denoted $\nabla f(x) \in \mathbb{R}^n$, such that for each direction $d \in \mathbb{R}^n$ and each scalar t,

$$f(x + td) - f(x) \;=\; t\langle \nabla f(x), d \rangle \;+\; e(t)$$

where the "error term" $e(t)$ is $o(t)$. (Note $\langle \nabla f(x), d \rangle$ is a scalar.) The column vector $\nabla f(x)$ is called the *gradient* of f at x.

To help you understand this definition of differentiability, recall that a function of *one* variable $f : \mathbb{R} \to \mathbb{R}$ is differentiable at the point x if and only if the limit

$$\lim_{t \to 0} \frac{f(x + t) - f(x)}{t}$$

exists (and is finite). When f is a function of *many* variables, $f : \mathbb{R}^n \to \mathbb{R}$, we can consider a one-dimensional "slice" of f, by considering at a point x the values of f in a single direction, $d \in \mathbb{R}^n$. We could imagine aligning an "x-axis" so that the point x is at the origin and the axis points in the direction d. Points along the axis can be described by $x + sd$ for all $s \in \mathbb{R}$. Then we can plot $\phi(s) = f(x + sd)$ as a function of s, obtaining a one-dimensional function, which is a slice of f in the direction d. Now the function ϕ is differentiable at $s = 0$ if and only if the limit

$$\lim_{t\to 0} \frac{\phi(s+t) - \phi(s)}{t} = \lim_{t\to 0} \frac{\phi(t) - \phi(0)}{t} = \lim_{t\to 0} \frac{f(x+td) - f(x)}{t}$$

exists. We actually call

$$f'(x;d) \stackrel{def}{=} \lim_{t\to 0} \frac{f(x+td) - f(x)}{t}$$

the *directional derivative of f at x in the direction d*. This is essentially the one-dimensional derivative of the slice of f along the direction d. To say that f is differentiable at x, we require that directional derivatives exist for all directions d. Now if there exists a vector $\nabla f(x)$ such that, for each direction d,

$$f(x+td) - f(x) = t\langle \nabla f(x), d\rangle + e(t)$$

where $e(t)$ is $o(t)$, then

$$\frac{f(x+td) - f(x)}{t} = \langle \nabla f(x), d\rangle + \frac{e(t)}{t},$$

so

$$\begin{aligned}
\lim_{t\to 0} \frac{f(x+td) - f(x)}{t} &= \lim_{t\to 0} \left(\langle \nabla f(x), d\rangle + \frac{e(t)}{t} \right) \\
&= \langle \nabla f(x), d\rangle + \lim_{t\to 0} \frac{e(t)}{t} \\
&= \langle \nabla f(x), d\rangle
\end{aligned}$$

since $\lim_{t\to 0} \frac{e(t)}{t} = 0$, as $e(t)$ is $o(t)$. Thus the directional derivative exists and is given by $\langle \nabla f(x), d\rangle$. Let e_i be the unit vector in direction i. Then the partial derivative in direction i is given by $\frac{\partial f(x)}{\partial x_i} = \langle \nabla f(x), e_i\rangle$. It follows that

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1} \ \cdots \ \frac{\partial f(x)}{\partial x_n} \right].$$

For a given point $y \in \mathbb{R}^n$, the gradient vector at that point, $\nabla f(y)$, is a vector normal (at right angles) to the tangent plane of the curve defined by the set of points $\{x \in \mathbb{R}^n \ : \ f(x) = f(y)\}$, at the point y.

We say f is C^1 or continuously differentiable if it is differentiable and the gradient function ∇f is continuous.

Example 1.2.1 Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$. Note that an equivalent expression for f is

$$f(x) = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$$

The derivative of f exists for all $x > 0$; it is $\nabla f(x) = 1$ for all $x > 0$. The derivative of f also exists for all $x < 0$; it is $\nabla f(x) = -1$ for all $x < 0$. However the derivative of f does *not* exist at $x = 0$, since

$$\lim_{t \to 0^+} \frac{f(0+t) - f(0)}{t} = \lim_{t \to 0^+} \frac{|t|}{t} = \lim_{t \to 0^+} \frac{t}{t} = 1$$

where $\lim\limits_{t \to 0^+}$ denotes the limit as t approaches 0 from above, whilst

$$\lim_{t \to 0^-} \frac{f(0+t) - f(0)}{t} = \lim_{t \to 0^-} \frac{|t|}{t} = \lim_{t \to 0^-} \frac{-t}{t} = -1,$$

where $\lim\limits_{t \to 0^-}$ denotes the limit as t approaches 0 from below, so

$$\lim_{t \to 0^+} \frac{f(0+t) - f(0)}{t} \neq \lim_{t \to 0^-} \frac{f(0+t) - f(0)}{t},$$

and hence the limit $\lim\limits_{t \to 0} \dfrac{f(0+t) - f(0)}{t}$ does not exist. (Alternatively, one can see that the gradient function

$$\nabla f(x) = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases}$$

must have a discontinuity at $x = 0$.) Thus f is not differentiable at $x = 0$, (its gradient function is not continuous), and so $f \notin \mathrm{C}^1$.

Example 1.2.2 Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = (x_+)^2$, where $x_+ = \max\{x, 0\}$. Note that an equivalent expression for f is

$$f(x) = \begin{cases} x^2, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The derivative of f exists for all $x > 0$; it is $\nabla f(x) = 2x$ for all $x > 0$. The derivative of f also exists for all $x < 0$; it is $\nabla f(x) = 0$ for all $x < 0$. The derivative of f also exists at $x = 0$, since

$$\lim_{t \to 0^+} \frac{f(0+t) - f(0)}{t} = \lim_{t \to 0^+} \frac{t^2}{t} = \lim_{t \to 0^+} t = 0$$

and
$$\lim_{t\to 0^-}\frac{f(0+t)-f(0)}{t} = \lim_{t\to 0^-}\frac{0}{t} = 0,$$
so
$$\lim_{t\to 0^+}\frac{f(0+t)-f(0)}{t} = 0 = \lim_{t\to 0^-}\frac{f(0+t)-f(0)}{t}$$
and hence the limit exists, i.e. $\nabla f(x) = \lim_{t\to 0}\dfrac{f(0+t)-f(0)}{t} = 0$. Thus the gradient function is
$$\nabla f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$
which is continuous. Thus f is differentiable, its gradient function is continuous, and so $f \in C^1$.

Hessians. We say f is twice differentiable if f is differentiable and ∇f is also differentiable, that is for each $x \in \mathbb{R}^n$, there is a matrix $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ such that for each $d \in \mathbb{R}^n$ and scalar t,
$$\nabla f(x+td) - \nabla f(x) \;=\; t\nabla^2 f(x)d + e(t)$$
where $e(t) = o(t)$. (Note $\nabla^2 f(x)d \in R^n$.)

The second derivative matrix $\nabla^2 f(x)$ is called the *Hessian* of f at x. It can be written using partial derivatives, namely the component in row i and column j of $\nabla^2 f(x)$ is
$$\nabla^2 f(x)_{ij} \;=\; \frac{\partial}{\partial x_j}\frac{\partial f(x)}{\partial x_i} \;=\; \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.$$

We say f is C^2 (twice continuously differentiable) if f is twice differentiable and the Hessian function $\nabla^2 f$ is continuous.

Exercise. Show that $f(x) = (x_+)^2$ given in Example 1.2.2 is not in C^2.

Exercise. Show that $f : \mathbb{R} \to \mathbb{R}$ given by
$$f(x) = \begin{cases} x^3 - 4x^2 + 5x - 2, & x \geq 1 \\ -(x-1)^2, & x < 1 \end{cases}$$
is in C^2. Give both $\nabla f(x)$ and $\nabla^2 f(x)$.

If f is C^2 then the Hessian of f at x is symmetric: $\nabla^2 f(x)_{ij} = \nabla^2 f(x)_{ji}$.

If $f(x) = a^T x$ for $a, x \in \mathbb{R}^n$, then $\nabla f(x) = a$ and $\nabla^2 f(x)$ is the $n \times n$ matrix of all zeroes.

If $f(x) = x^T Bx$ for $x \in \mathbb{R}^n$, $B \in \mathbb{R}^{n\times n}$, B symmetric, then $\nabla f(x) = 2Bx$ and $\nabla^2 f(x) = 2B$.

Differentiability. We say $f : \mathbb{R} \to \mathbb{R}$ is C^k (k-times continuously differentiable) if f is differentiable k times, and the derivative functions $\nabla^j f(x)$ are continuous for all $j = 1, \ldots, k$. (Note $\nabla^1 f(x) = \nabla f(x)$.) We say f is C^∞ (infinitely continuously differentiable) if $f, \nabla f, \nabla^2 f, \nabla^3 f, \ldots$ are all differentiable, and the derivative functions $\nabla^j f(x)$ are continuous for all $j = 1, 2, 3 \ldots$. In some cases, we will talk about differentiability over a specific domain, e.g. over $x > 0$.

Note that for functions of many variables, derivatives of order higher than 2 are difficult to define, so we omit discussion of them, and will consider higher derivatives only for functions of a single variable.

All polynomial functions are in C^∞.

The function $\log x$ over $x > 0$ is in C^∞.

The function $\frac{1}{x}$ over $x > 0$ is in C^∞.

Exercise. What is the largest value of k such that $f(x) = (|x|)^5$ is in C^k?

Taylor's theorem. For $x, d \in \mathbb{R}^n$ and $t \in \mathbb{R}$,

$$\begin{aligned}
f(x+td) &= f(x) + t\langle \nabla f(x), d\rangle + o(t) \\
&\quad \text{if } f \text{ is } C^1 \\
&= f(x) + t\langle \nabla f(x), d\rangle + \frac{t^2}{2}\langle d, \nabla^2 f(x) d\rangle + o(t^2) \\
&\quad \text{if } f \text{ is } C^2.
\end{aligned}$$

Convexity. We say f is *convex* if for each $x, y \in \mathbb{R}^n$ and $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Suppose f is C^2. Then f is convex if and only if $\nabla^2 f(x)$ is positive semi-definite for every $x \in \mathbb{R}^n$. Also, if f is convex then $\nabla^2 f(x)$ is invertible if and only if $\nabla^2 f(x)$ is positive definite.

Quadratic Functions. A quadratic function is any function of the form

$$f(x) = \alpha + \langle c, x\rangle + \tfrac{1}{2}\langle x, Bx\rangle = \alpha + c^T x + \tfrac{1}{2} x^T B x, \qquad (1.1)$$

where $\alpha \in \mathbb{R}$, $c \in \mathbb{R}^n$, and B is a symmetric matrix in $\mathbb{R}^{n\times n}$. Such a function is C^2:

$$\begin{aligned}
\nabla f(x) &= c + \tfrac{1}{2}Bx + \tfrac{1}{2}B^T x = c + Bx \text{ (by symmetry)}, \\
\nabla^2 f(x) &= B.
\end{aligned}$$

A function $f : \mathbb{R}^3 \to \mathbb{R}$ is quadratic if and only if it has the form

$$\begin{aligned} f(x) &= \alpha + c_1x_1 + c_2x_2 + c_3x_3 \\ &\quad + B_{12}x_1x_2 + B_{23}x_2x_3 + B_{13}x_1x_3 \\ &\quad + \tfrac{1}{2}B_{11}x_1^2 + \tfrac{1}{2}B_{22}x_2^2 + \tfrac{1}{2}B_{33}x_3^2 \end{aligned}$$

for some constants α, c_1, c_2, c_3, and $B_{12}, B_{23}, B_{13}, B_{11}, B_{22}, B_{33}$.

Note that such a function is given by (1.1) with $c = (c_1, c_2, c_3)$ and

$$B = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}.$$

In particular, you can write down B simply by looking at the coefficients of the degree two terms x_1^2, x_1x_2 etc.

Note that

$$\nabla f(x) = c + Bx$$

and

$$\nabla^2 f(x) = B.$$

Thus a quadratic function f is convex if and only if B is positive semidefinite.

Unconstrained Optimization of Functions of a Single Variable. To optimize a function $f : \mathbb{R} \to \mathbb{R}$, over a domain $x \in [a, b]$, where $f \in C^1$, (and possibly $a = \infty$ or $b = \infty$), is often easy. In particular, it is easy in the case that the equation $\nabla f(x) = 0$ can be solved analytically to determine all roots. In such cases the optimal value must occur at one of these roots, or at one of the (finite) endpoints of the domain. The sign of the derivative can be used to rule out some roots. The optimal solution can be determined by simply evaluating the function at each of these points.

In the case that the equation $\nabla f(x) = 0$ *cannot* be solved analytically, numerical optimization techniques are required. A good part of this course will discuss such techniques.

Example 1.2.3 To find the minimum value of $f(x) = 2x^3 - 3x^2 - 12x + 30$ over the domain $[-2, \infty)$, that is, solve the optimization problem

$$\min_{x \geq -2} f(x) = 2x^3 - 3x^2 - 12x + 17,$$

we observe $\nabla f(x) = 6x^2 - 6x - 12 = 6(x-2)(x+1)$. Now finding the roots of $\nabla f(x) = 0$ is easy: clearly these are 2 and -1. Since $f \in C^1$, we know that the sign of $\nabla f(x)$ must be the same for all $x < -1$, all x with $-1 < x < 2$ and for all $x > 2$. In fact, $\nabla f(x) > 0$ for all $x > -1$ and for all $x > 2$, and $\nabla f(x) < 0$ for all x with $-1 < x < 2$. Since the sign changes from positive to negative as x increases through $x = -1$, this is a local maximum, and can be ignored. However the sign changes from negative to positive as x increases through $x = 2$, so this root is a local minimum and needs to be considered. As $\nabla f(x) > 0$ for all $x > 2$, f must be increasing for $x > 2$, so any minimum must occur in the domain $x \leq 2$. We thus need only check the endpoint of the given domain, $x = -2$, and the local minimum $x = 1$. The value of the former is $f(-2) = -16 - 12 + 24 + 17 = 13$ and the value of the latter is $f(1) = 2-3-12+17 = 4$. Thus the optimal solution is $x = 1$, with value $f(1) = 4$.

For $q(x) = cx^2 + dx + k$ a quadratic function with $c > 0$, the solution of

$$\min_{x \geq a} q(x) = cx^2 + dx + k$$

is given by $x = \max\{a, -\frac{d}{2c}\}$. The solution of

$$\min_{x \leq b} q(x) = cx^2 + dx + k$$

is given by $x = \min\{b, -\frac{d}{2c}\}$. The solution of

$$\min_{a \leq x \leq b} q(x) = cx^2 + dx + k$$

is given by

$$x = \begin{cases} -\frac{d}{2c}, & a \leq -\frac{d}{2c} \leq b \\ a, & -\frac{d}{2c} < a \\ b, & -\frac{d}{2c} > b. \end{cases}$$