ORII: Optimization

Simon Abelard

January 2026

The faces of optimization

Remember: many types of problems, many algorithms!

Optimization can be:

- Continuous (LP) or discrete (ILP)
- Linear (LP) or non-linear (convexity often helps)
- Constrained (LP) or unconstrained (differential conditions)
- Global (LP) or local (derivatives are useful)
- Deterministic or stochastic (LP could be both)
- First-order (gradient descent) or higher-order
- And many more (genetic algorithms, deep learning, ...)

Syllabus:

- Extrema of nonlinear functions
- Detection and characterization of extrema
- Algorithms: Gradient-descent, Newton's method
- Convex Optimization
- Lagrange/Karush-Kuhn-Tucker conditions

Grading process: one exam (similar to OR I).
January 28^{th} at 9:30 pm, duration 2 hours.
Material allowed: one (nominative) cheat sheet.

Basics of calculus: multivariate derivatives

Let $f(x_1, \ldots, x_n)$ be a $\mathcal{C}^2$ function from $\mathbb{R}^n$ to $\mathbb{R}$.

The **gradient** of f at x is the vector $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)$. Geometrically the gradient is the direction of steepest ascent, hence its use in optimization.

The Hessian of f at x is the symmetric matrix:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \ldots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial^2 x_1} & \ldots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}$$

Both the gradient and Hessian are used for optimality criteria, i.e. detecting and deciding whether extrema are maxima or minima.

Extrema and optimization

Global extrema

Let $f : C \subset \mathbb{R}^n \to \mathbb{R}$, we say that x_0 is a **global minimum** (resp. maximum) if $\forall x \in C, f(x_0) \leq f(x)$ (resp. $f(x_0) \geq f(x)$).

Local extrema

Let $f : C \subset \mathbb{R}^n \to \mathbb{R}$, we say that x_0 is a **local minimum** (resp. maximum) if there exists a $D > 0$ such that for all x satisfying $||x - x_0|| \leq D$, we have $f(x_0) \leq f(x)$ (resp. $f(x_0) \geq f(x)$).

Extrema and optimization

Global extrema

Let $f : C \subset \mathbb{R}^n \to \mathbb{R}$, we say that x_0 is a **global minimum** (resp. maximum) if $\forall x \in C, f(x_0) \leq f(x)$ (resp. $f(x_0) \geq f(x)$).

Local extrema

Let $f : C \subset \mathbb{R}^n \to \mathbb{R}$, we say that x_0 is a **local minimum** (resp. maximum) if there exists a $D > 0$ such that for all x satisfying $||x - x_0|| \leq D$, we have $f(x_0) \leq f(x)$ (resp. $f(x_0) \geq f(x)$).

Unconstrained optimization: finding a global extremum.
Constrained optimization: finding a global extremum over a subset defined by constraints (i.e. among feasible solutions).

Extrema and optimization

Global extrema

Let $f : C \subset \mathbb{R}^n \to \mathbb{R}$, we say that x_0 is a **global minimum** (resp. maximum) if $\forall x \in C, f(x_0) \leq f(x)$ (resp. $f(x_0) \geq f(x)$).

Local extrema

Let $f : C \subset \mathbb{R}^n \to \mathbb{R}$, we say that x_0 is a **local minimum** (resp. maximum) if there exists a $D > 0$ such that for all x satisfying $||x - x_0|| \leq D$, we have $f(x_0) \leq f(x)$ (resp. $f(x_0) \geq f(x)$).

Unconstrained optimization: finding a global extremum.
Constrained optimization: finding a global extremum over a subset defined by constraints (i.e. among feasible solutions).

Questions of this course:

- Differential calculus $\rightsquigarrow$ local extrema **when no constraints**
- Existence of global extrema? (convexity)
- Detecting optimality under constraints? (Lagrange, KKT)

Unconstrained optimization: local extrema

Characterizing local extrema

Let f be a $\mathcal{C}^2$ function. The following holds:

- If x is a local extrema, then $\nabla f(x) = 0$ (x is a critical point)
- If the Hessian matrix $\nabla^2 f(x)$
  - is positive definite, then x is a local minimum
  - is negative definite, then x is a local maximum
  - has eigenvalues of both signs, then x is a saddle point

Unconstrained optimization: local extrema

Characterizing local extrema

Let f be a $\mathcal{C}^2$ function. The following holds:

- If x is a local extrema, then $\nabla f(x) = 0$ (x is a critical point)
- If the Hessian matrix $\nabla^2 f(x)$
  - is positive definite, then x is a local minimum
  - is negative definite, then x is a local maximum
  - has eigenvalues of both signs, then x is a saddle point

Idea: using multivariate Taylor series expansion at x, we have

$$f(x + \Delta x) = f(x) + (\Delta x)^T \nabla^2 f(x) \Delta x = f(x) + q(\Delta x)$$

So if $q(\Delta x)$ is positive (resp negative) in a neighborhood of x, we indeed deduce that x is a minimum (resp. maximum)

Examples

Find the extrema of the function $f(x, y) = x^2 + y^2 - 4x - 6y + 9$.

Examples

Find the extrema of the function $f(x, y) = x^2 + y^2 - 4x - 6y + 9$.

$\nabla f(x) = (2x - 4, 2y - 6)$ so $(2, 3)$ is the only critical point.
Compute the Hessian matrix $\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. It has one positive eigenvalue of multiplicity 2 hence it is positive definite.
Conclusion: $(x, y) = (2, 3)$ is a **local** minimum of f.

Examples

Find the extrema of the function $f(x, y) = x^2 + y^2 - 4x - 6y + 9$.

$\nabla f(x) = (2x - 4, 2y - 6)$ so $(2, 3)$ is the only critical point.

Compute the Hessian matrix $\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. It has one positive eigenvalue of multiplicity 2 hence it is positive definite.

Conclusion: $(x, y) = (2, 3)$ is a **local** minimum of f.

Find the extrema of the function $f(x, y) = x^3 - 3xy^2$

Examples

Find the extrema of the function $f(x,y) = x^2 + y^2 - 4x - 6y + 9$.

$\nabla f(x) = (2x-4, 2y-6)$ so $(2,3)$ is the only critical point.
Compute the Hessian matrix $\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. It has one positive eigenvalue of multiplicity 2 hence it is positive definite.
Conclusion: $(x,y) = (2,3)$ is a **local** minimum of f.

Find the extrema of the function $f(x,y) = x^3 - 3xy^2$

$\nabla f(x) = (3x^2 - 3y^2, -6xy)$, so $(0,0)$ is the only critical point.
Compute the Hessian matrix $\nabla^2 f(x) = \begin{pmatrix} 6x & -6y \\ -6y & -6x \end{pmatrix}$.
At $(0,0)$, the Hessian is the zero matrix so it is neither positive nor negative definite.
Conclusion: we cannot conclude.

Counter-examples

- **If $\nabla^2 f(x)$ is only positive semi-definite:** x is not always an extremum.
  See $f(x) = x^3$, $x = 0$ is a critical point and the Hessian ($f''(0)$) is zero. It is clear that $x = 0$ is not an extremum.
- **But if $\nabla^2 f(x)$ is not positive definite, it does not mean that x cannot be a minimum!**
  See $f(x) = x^4$, $x = 0$ is a critical point and the Hessian ($f''(0)$) is zero. But $x = 0$ is a (global) minimum of f.

Conclusion: be very careful with necessary/sufficient conditions.

An extremum is a critical point but the converse is not rue.
When the theorem above does not say anything (e.g. Hessian is zero), anything can happen (minimum, maximum or neither).

Newton's method for local extrema

The algorithm:

- Newton's method: find roots of functions
- Works for several variables: $\nabla g(x_n)^T(x_n - x_{n+1}) = -g(x_n)$
- Use it with $g = f'$ to find critical points of f
- Look at the Hessian of f to conclude (or not)

Newton's method for local extrema

The algorithm:

- Newton's method: find roots of functions
- Works for several variables: $\nabla g(x_n)^T(x_n - x_{n+1}) = -g(x_n)$
- Use it with $g = f'$ to find critical points of f
- Look at the Hessian of f to conclude (or not)

The limits

- Newton's method does not always work
- Not always sufficient to conclude (Hessian)
- Even so, extrema are local and not global

Gradient descent

Let f be a differentiable function that we want to minimize.

Idea: $-\nabla f(x)$ is the direction of steepest descent at point x, so let's follow this direction to find a local minimum.

Recursion: $x_{n+1} = x_n - \eta_n \nabla f(x_n)$

Gradient descent

Let f be a differentiable function that we want to minimize.

Idea: $-\nabla f(x)$ is the direction of steepest descent at point x, so let's follow this direction to find a local minimum.

Recursion: $x_{n+1} = x_n - \eta_n \nabla f(x_n)$

Learning rate: $\eta_n =$ how far we go in this direction.
Neither too small (slow convergence) nor too large (overshoot).

Stopping criterion: when $||x_{n+1} - x_n||$ is small enough.

Gradient descent

Let f be a differentiable function that we want to minimize.

Idea: $-\nabla f(x)$ is the direction of steepest descent at point x, so let's follow this direction to find a local minimum.

Recursion: $x_{n+1} = x_n - \eta_n \nabla f(x_n)$

Learning rate: η_n = how far we go in this direction.
Neither too small (slow convergence) nor too large (overshoot).

Stopping criterion: when $||x_{n+1} - x_n||$ is small enough.

Choice of η: crucial but complicated, can be

- deterministic or stochastic
- based on the Hessian (if f is $\mathcal{C}^2$)
- optimized in certain cases (f convex, Lipschitz, linear)

Limits: local algorithm, success and speed is not guaranteed

Convexity

Can we look for easier instances of the problem?

Remember linear programming:

- Feasible region is always a polyhedron
- If it exists, optimal solution is a vertex
- Optimality conditions (complementary slackness)

This is not true if the problem is no longer linear, but when it is convex we have:

- The optimal set is a convex set
- Every local minimum is global
- Lagrange multipliers (generalize complementary slackness)

Convexity

Definition (convex function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if for any distinct x and y and any $\lambda \in]0, 1[$ we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

If the inequality is strict, we say that f is **strictly convex**.

A function g is called **concave** if $-g$ is convex.

Convexity

Definition (convex function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if for any distinct x and y and any $\lambda \in]0, 1[$ we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

If the inequality is strict, we say that f is **strictly convex**.

A function g is called **concave** if $-g$ is convex.

Examples:

- Affine functions are convex
- Norms are convex (triangular inequality)
- Quadratic functions are convex

Proving that a function is convex

Option 1: using known facts (exercise: prove these facts)

- f is affine or f is a norm
- f is a positive linear combination of convex functions
- $f = g \circ h$ with g convex and h affine

Proving that a function is convex

Option 1: using known facts (exercise: prove these facts)

- f is affine or f is a norm
- f is a positive linear combination of convex functions
- $f = g \circ h$ with g convex and h affine

Option 2: using the gradient
A $\mathcal{C}^1$ function f is convex over C if and only if for any $x, y \in C$ we have $(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$.

Proving that a function is convex

Option 1: using known facts (exercise: prove these facts)

- f is affine or f is a norm
- f is a positive linear combination of convex functions
- $f = g \circ h$ with g convex and h affine

Option 2: using the gradient
A $\mathcal{C}^1$ function f is convex over C if and only if for any $x, y \in C$ we have $(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$.

Option 3: using the Hessian
A $\mathcal{C}^2$ function f is convex over C if and only if for any $x \in C$, $\nabla^2 f(x)$ is positive semi-definite.

Convex optimization

Convex functions are easier for optimization because:

Theorem: stationarity implies global minimality

If f is a convex $\mathcal{C}^1$ function over a **convex** set C such that $x^* \in C$ is a stationary point of f (i.e. $\nabla f(x^*) = 0$) then x^* is the global minimizer of f over C (i.e. $\forall x \in C, f(x) \geq f(x^*)$).

Convex optimization

Convex functions are easier for optimization because:

Theorem: stationarity implies global minimality

If f is a convex $\mathcal{C}^1$ function over a **convex** set C such that $x^* \in C$ is a stationary point of f (i.e. $\nabla f(x^*) = 0$) then x^* is the global minimizer of f over C (i.e. $\forall x \in C, f(x) \geq f(x^*)$).

Theorem: maximization over a compact

If f is convex over a **nonempty, convex and compact** set C then f admits at least one maximizer which is an extreme point of C.

Examples: unconstrained optimization

What is $\min_{x,y} f(x,y)$, for $f(x,y) = 4x^2 + 2y^2 - xy + 3x + 1$?

Examples: unconstrained optimization

What is $\min_{x,y} f(x,y)$, for $f(x,y) = 4x^2 + 2y^2 - xy + 3x + 1$?

- f is (strictly) convex because $\nabla^2 f(x,y) = \begin{pmatrix} 8 & -1 \\ -1 & 4 \end{pmatrix}$ is definite positive (its determinant and trace are both positive).

Examples: unconstrained optimization

What is $\min_{x,y} f(x,y)$, for $f(x,y) = 4x^2 + 2y^2 - xy + 3x + 1$?

- f is (strictly) convex because $\nabla^2 f(x,y) = \begin{pmatrix} 8 & -1 \\ -1 & 4 \end{pmatrix}$ is definite positive (its determinant and trace are both positive).
- $\nabla f(x,y) = (8x - y + 3, 4y - x)^T$ hence f has a single critical point in $(x_0, y_0) = (-12/31, -3/31)$.

Examples: unconstrained optimization

What is $\min_{x,y} f(x,y)$, for $f(x,y) = 4x^2 + 2y^2 - xy + 3x + 1$?

- f is (strictly) convex because $\nabla^2 f(x,y) = \begin{pmatrix} 8 & -1 \\ -1 & 4 \end{pmatrix}$ is definite positive (its determinant and trace are both positive).
- $\nabla f(x,y) = (8x - y + 3, 4y - x)^T$ hence f has a single critical point in $(x_0, y_0) = (-12/31, -3/31)$.
- By strict convexity, (x_0, y_0) is the global minimum of f.

Examples: unconstrained optimization

What is $\min_{x,y} f(x,y)$, for $f(x,y) = 4x^2 + 2y^2 - xy + 3x + 1$?

- f is (strictly) convex because $\nabla^2 f(x,y) = \begin{pmatrix} 8 & -1 \\ -1 & 4 \end{pmatrix}$ is definite positive (its determinant and trace are both positive).
- $\nabla f(x,y) = (8x - y + 3, 4y - x)^T$ hence f has a single critical point in $(x_0, y_0) = (-12/31, -3/31)$.
- By strict convexity, (x_0, y_0) is the global minimum of f.
- The answer is thus $f(x_0, y_0) = \frac{475}{961} \simeq 0.49$.

Examples: unconstrained optimization

What is $\min_{x,y} f(x,y)$, for $f(x,y) = 4x^2 + 2y^2 - xy + 3x + 1$?

- f is (strictly) convex because $\nabla^2 f(x,y) = \begin{pmatrix} 8 & -1 \\ -1 & 4 \end{pmatrix}$ is definite positive (its determinant and trace are both positive).
- $\nabla f(x,y) = (8x - y + 3, 4y - x)^T$ hence f has a single critical point in $(x_0, y_0) = (-12/31, -3/31)$.
- By strict convexity, (x_0, y_0) is the global minimum of f.
- The answer is thus $f(x_0, y_0) = \frac{475}{961} \simeq 0.49$.

What is $\max_{(x,y) \in B(0,1)} f(x,y)$?

The unit ball $B = B(0,1)$ is compact and f is convex on B so f has a maximum which is reached on $S(0,1)$ the unit sphere.
We can thus add the constraint $x^2 + y^2 = 1$ to the problem.

Constrained optimization

Non-linear Problem:

$$\begin{cases} \min_{x \in C} f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

This problem is often denoted by $\mathcal{P}_{EI}$, and by $\mathcal{P}_E$ if there is no inequality constraint

Technically, f, g and h can be any functions but very little can be said.

Constrained optimization

Non-linear Problem:

$$\begin{cases} \min_{x \in C} f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

This problem is often denoted by $\mathcal{P}_{EI}$, and by $\mathcal{P}_E$ if there is no inequality constraint

Technically, f, g and h can be any functions but very little can be said.

Definition: convex problem

The problem $\mathcal{P}_{EI}$ is **convex** if:

- The objective/criteria f is convex on the feasible set
- The function g is convex on the feasible set
- The function h is **linear** (Warning, linear not just convex!)

Remark: $\mathcal{P}_E$ is convex when the first two points are satisfied ($h = 0$ is convex and $h(x) \leq 0$ holds for any x)

Optimality conditions, the general case

Consider the NL problem:

$$\begin{cases} \min\limits_{x \in C} f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

We assume that f, g and h are $\mathcal{C}^1$ functions over C.

Karush-Kuhn-Tucker (KKT) conditions: if x^* is an optimal solution, then there is a linear relationship between the gradients of f, g and h.

Not very useful because the converse is false, so finding a point satisfying KKT conditions will not give a minimum.

With stronger assumptions (convexity in particular), the we can have a stronger and more useful version of KKT conditions.

Optimality conditions, the convex case

Consider the NL problem:

$$\begin{cases} \min_{x \in C} f(x) \\ g_i(x) \leq 0 \\ h_j(x) = 0 \end{cases}$$

We assume that f, the g_i's and the h_j's are $\mathcal{C}^1$, and that the problem is **convex**

Theorem (Karush-Kuhn-Tucker)

If the constraints are **qualified**, then x^* is a global minimum of the problem if and only if there exists $(\lambda_i) \in \mathbb{R}_+^m$ and $(\mu_j) \in \mathbb{R}^k$ such that

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^k \mu_j \nabla h_j(x^*) = 0, \\ \forall i \leq m, \quad g_i(x) \leq 0, \\ \forall i \leq m, \quad \lambda_i \geq 0 \text{ and } \lambda_i g_i(x^*) = 0, \\ \forall j \leq k, \quad h_j(x) = 0. \end{cases}$$

A bit more on Karush-Kuhn-Tucker

The constraints can be split into:

- Stationarity (nullity of the gradient)
- Feasibility ($g_i(x^*) \leq 0$, $h_i(x^*) = 0$)
- Positivity of dual variables ($\lambda_i \geq 0$)
- **Complementary slackness** $\lambda_i g_i(x^*) = 0$

What about the "qualified" part? It is technical so let us just use:

Proposition (Slater's conditions)

If there exists in α such that $h_j(\alpha) = 0$ for all $j \leq k$ and $g_i(\alpha) < 0$ for all $i \leq m$ then the constraints are qualified.

Proposition (LICQ)

Let B be the family containing the vectors $(\nabla h_j(x^*))_{j \leq k}$ and the vectors $\nabla g_i(x^*)$, for i such that g_i is active in x^* (i.e. $\lambda_i \neq 0$). If B is a free family, then the constraints are qualified.

Counter-example

What happens when qualification fails?
An optimal point does not necessarily satisfies KKT conditions.

Example:

$$\begin{cases} \min\limits_{x_1, x_2 \in \mathbb{R}} x_2 \\ (x_1 - 1)^2 + x_2^2 \leq 1 \\ (x_1 + 1)^2 + x_2^2 \leq 1 \end{cases}$$

The only feasible point is $x^* = (0, 0)$ so it is obviously optimal.
$\nabla f(x^*) = (0, 1)^T$, $\nabla g_1(x^*) = (-2, 0)^T$ and $\nabla g_2(x^*) = (2, 0)^T$.
Thus, $\nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) = (2(\lambda_2 - \lambda_1), 1)^T$.

No matter the choice of (λ_1, λ_2) it will never be 0.

Therefore x^* is optimal and does not satisfy KKT conditions

Example: writing KKT conditions

Consider the problem:

$$\begin{cases} \min\limits_{x,y \in \mathbb{R}} f(x,y) = x^2 + y^2 \\ x \geq 0 \\ y \geq 0 \\ x + 2y = 4 \end{cases}$$

Example: writing KKT conditions

Consider the problem:

$$\begin{cases} \min\limits_{x,y\in\mathbb{R}} f(x,y) = x^2 + y^2 \\ x \geq 0 \\ y \geq 0 \\ x + 2y = 4 \end{cases}$$

In standard form:

$$\begin{cases} \min\limits_{x,y\in\mathbb{R}} x^2 + y^2 \\ -x \leq 0 \\ -y \leq 0 \\ x + 2y = 4 \end{cases}$$

Example: writing KKT conditions

Consider the problem:

$$\begin{cases} \min\limits_{x,y\in\mathbb{R}} f(x,y) = x^2 + y^2 \\ x \geq 0 \\ y \geq 0 \\ x + 2y = 4 \end{cases}$$

In standard form:

$$\begin{cases} \min\limits_{x,y\in\mathbb{R}} x^2 + y^2 \\ -x \leq 0 \\ -y \leq 0 \\ x + 2y = 4 \end{cases}$$

The problem is convex so f has a unique global minimum.
Slater's condition (at $\alpha = (2,1)$) $\implies$ constraints are qualified.
By reciprocal KKT conditions, we will find the optimum by solving:

Example: writing KKT conditions

Consider the problem:

$$\begin{cases} \min_{x,y\in\mathbb{R}} f(x,y) = x^2 + y^2 \\ x \geq 0 \\ y \geq 0 \\ x + 2y = 4 \end{cases}$$

In standard form:

$$\begin{cases} \min_{x,y\in\mathbb{R}} x^2 + y^2 \\ -x \leq 0 \\ -y \leq 0 \\ x + 2y = 4 \end{cases}$$

The problem is convex so f has a unique global minimum.
Slater's condition (at $\alpha = (2,1)$) $\implies$ constraints are qualified.
By reciprocal KKT conditions, we will find the optimum by solving:

$$\begin{cases} 2x - \lambda_1 + \mu = 0 \\ 2y - \lambda_2 + 2\mu = 0 \\ -x \leq 0, \quad -y \leq 0 \\ x + 2y = 4 \\ \lambda_1, \lambda_2 \geq 0 \\ \lambda_1 x = \lambda_2 y = 0 \end{cases}$$

Example: solving KKT conditions

$$\begin{cases} 2x - \lambda_1 + \mu = 0 \\ 2y - \lambda_2 + 2\mu = 0 \\ -x \leq 0, \quad -y \leq 0 \\ x + 2y = 4 \\ \lambda_1, \lambda_2 \geq 0 \\ \lambda_1 x = \lambda_2 y = 0 \end{cases}$$

Usually we solve it systematically assuming positivity of multipliers. If $\lambda_1 > 0$, then $x = 0$ by complementary slackness.

Example: solving KKT conditions

$$
\begin{cases}
2x - \lambda_1 + \mu = 0 \\
2y - \lambda_2 + 2\mu = 0 \\
-x \le 0, \quad -y \le 0 \\
x + 2y = 4 \\
\lambda_1, \lambda_2 \ge 0 \\
\lambda_1 x = \lambda_2 y = 0
\end{cases}
$$

Usually we solve it systematically assuming positivity of multipliers. If $\lambda_1 > 0$, then $x = 0$ by complementary slackness.
By the equality, $y = 2$ and so $\lambda_2 = 0$.
Gradient conditions: $\mu = \lambda_1 = -y = -2$, which is impossible since $\lambda_1 > 0$.
Conclusion: no solution for $\lambda_1 > 0$.

Example: solving KKT conditions

$$\begin{cases} 2x - \lambda_1 + \mu = 0 \\ 2y - \lambda_2 + 2\mu = 0 \\ -x \leq 0, \quad -y \leq 0 \\ x + 2y = 4 \\ \lambda_1, \lambda_2 \geq 0 \\ \lambda_1 x = \lambda_2 y = 0 \end{cases}$$

Usually we solve it systematically assuming positivity of multipliers. If $\lambda_1 > 0$, then $x = 0$ by complementary slackness.
By the equality, $y = 2$ and so $\lambda_2 = 0$.
Gradient conditions: $\mu = \lambda_1 = -y = -2$, which is impossible since $\lambda_1 > 0$.
Conclusion: no solution for $\lambda_1 > 0$.

Assuming $\lambda_2 > 0$, $y = 0$ so $x = 4$ and $\lambda_1 = 0$. The gradient conditions yield $\lambda_2 = -16 < 0$ which is a contradiction.

Example: solving KKT conditions

$$\begin{cases} 2x - \lambda_1 + \mu = 0 \\ 2y - \lambda_2 + 2\mu = 0 \\ -x \leq 0, \quad -y \leq 0 \\ x + 2y = 4 \\ \lambda_1, \lambda_2 \geq 0 \\ \lambda_1 x = \lambda_2 y = 0 \end{cases}$$

Usually we solve it systematically assuming positivity of multipliers. If $\lambda_1 > 0$, then $x = 0$ by complementary slackness.
By the equality, $y = 2$ and so $\lambda_2 = 0$.
Gradient conditions: $\mu = \lambda_1 = -y = -2$, which is impossible since $\lambda_1 > 0$.
Conclusion: no solution for $\lambda_1 > 0$.

Assuming $\lambda_2 > 0$, $y = 0$ so $x = 4$ and $\lambda_1 = 0$. The gradient conditions yield $\lambda_2 = -16 < 0$ which is a contradiction.

Remaining option: $\lambda_1 = \lambda_2 = 0$ so by the gradient conditions $y = -\mu$ and $2x = y$. Using the equality condition, $x = \frac{4}{5}$ so $y = \frac{8}{5}$. This satisfies all equations including $x \geq 0$ and $y \geq 0$.

Conclusion: the optimal solution of the problem is $\left(\frac{4}{5}, \frac{8}{5}\right)$, at which f reaches its global minimum equal to $\frac{16}{5}$.