香港中文大學
The Chinese University of Hong Kong

Chapter 9
The Blahut-Arimoto Algorithms

The Chinese University of Hong Kong

Single-Letter Characterization

Single-Letter Characterization

- For a DMC $p(y|x)$, the capacity

$$C = \max_{r(x)} I(X;Y),$$

where $r(x)$ is the input distribution, gives the maximum asymptotically achievable rate for reliable communication as the blocklength $n \to \infty$.

Single-Letter Characterization

- For a DMC $p(y|x)$, the capacity

$$C = \max_{r(x)} I(X;Y),$$

  where $r(x)$ is the input distribution, gives the maximum asymptotically achievable rate for reliable communication as the blocklength $n \to \infty$.

- This characterization of C, in the form of an optimization problem, is called a single-letter characterization because it involves only $p(y|x)$ but not n.

Single-Letter Characterization

- For a DMC $p(y|x)$, the capacity

$$C = \max_{r(x)} I(X;Y),$$

where $r(x)$ is the input distribution, gives the maximum asymptotically achievable rate for reliable communication as the blocklength $n \to \infty$.

- This characterization of C, in the form of an optimization problem, is called a single-letter characterization because it involves only $p(y|x)$ but not n.

- Similarly, the rate-distortion function

$$R(D) = \min_{Q(\hat{x}|x): Ed(X,\hat{X}) \le D} I(X;\hat{X})$$

for an i.i.d. information source $\{X_k\}$ is a single-letter characterization.

Numerical Methods

Numerical Methods

- When the alphabets are finite, C and $R(D)$ are given as solutions of finite-dimensional optimization problems.

Numerical Methods

- When the alphabets are finite, C and $R(D)$ are given as solutions of finite-dimensional optimization problems.
- However, these quantities cannot be expressed in closed-forms except for very special cases.

Numerical Methods

- When the alphabets are finite, C and $R(D)$ are given as solutions of finite-dimensional optimization problems.
- However, these quantities cannot be expressed in closed-forms except for very special cases.
- Even computing these quantities is not straightforward because the associated optimization problems are nonlinear.

Numerical Methods

- When the alphabets are finite, C and $R(D)$ are given as solutions of finite-dimensional optimization problems.
- However, these quantities cannot be expressed in closed-forms except for very special cases.
- Even computing these quantities is not straightforward because the associated optimization problems are nonlinear.
- So we have to resort to numerical methods.

Numerical Methods

- When the alphabets are finite, C and $R(D)$ are given as solutions of finite-dimensional optimization problems.
- However, these quantities cannot be expressed in closed-forms except for very special cases.
- Even computing these quantities is not straightforward because the associated optimization problems are nonlinear.
- So we have to resort to numerical methods.
- The Blahut-Arimoto (BA) algorithms are iterative algorithms devised for this purpose.

In this chapter

In this chapter

- A general alternating optimization algorithm

In this chapter

- A general alternating optimization algorithm
- The Blahut-Arimoto algorithms for computing
  1. the channel capacity C
  2. the rate-distortion function $R(D)$

In this chapter

- A general alternating optimization algorithm
- The Blahut-Arimoto algorithms for computing
  1. the channel capacity C
  2. the rate-distortion function $R(D)$
- Convergence of the alternating optimization algorithms

香港中文大學
The Chinese University of Hong Kong

9.1 Alternating Optimization

A Double Supremum

A Double Supremum

Consider the double supremum

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

A Double Supremum

Consider the double supremum

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

- A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

A Double Supremum

Consider the double supremum

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

- A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
- $f : A_1 \times A_2 \to \Re$ is bounded from above, such that
  - f is continuous and has continuous partial derivatives on $A_1 \times A_2$

A Double Supremum

Consider the double supremum

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

- A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
- $f : A_1 \times A_2 \to \Re$ is bounded from above, such that
  - f is continuous and has continuous partial derivatives on $A_1 \times A_2$
  - For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

A Double Supremum

Consider the double supremum

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

- A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
- $f : A_1 \times A_2 \to \Re$ is bounded from above, such that
  - f is continuous and has continuous partial derivatives on $A_1 \times A_2$
  - For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

    $$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

    and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

    $$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

A Double Supremum

- Let $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ and $A = A_1 \times A_2$. Then the double supremum can be written as

$$\sup_{\mathbf{u} \in A} f(\mathbf{u}).$$

A Double Supremum

- Let $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ and $A = A_1 \times A_2$. Then the double supremum can be written as

$$\sup_{\mathbf{u} \in A} f(\mathbf{u}).$$

- In other words, the supremum of f is taken over a subset of $\Re^{n_1+n_2}$ which is equal to the Cartesian product of two convex subsets of $\Re^{n_1}$ and $\Re^{n_2}$.

A Double Supremum

- Let $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ and $A = A_1 \times A_2$. Then the double supremum can be written as

$$\sup_{\mathbf{u} \in A} f(\mathbf{u}).$$

- In other words, the supremum of f is taken over a subset of $\Re^{n_1+n_2}$ which is equal to the Cartesian product of two convex subsets of $\Re^{n_1}$ and $\Re^{n_2}$.

- Let

$$f^* = \sup_{\mathbf{u} \in A} f(\mathbf{u}).$$

An Alternating Optimization

An Alternating Optimization

- Let $\mathbf{u}^{(k)} = (\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)})$ for $k \geq 0$, defined as follows.

An Alternating Optimization

- Let $\mathbf{u}^{(k)} = (\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)})$ for $k \geq 0$, defined as follows.
- Let $\mathbf{u}_1^{(0)}$ be an arbitrarily chosen vector in A_1, and let $\mathbf{u}_2^{(0)} = c_2(\mathbf{u}_1^{(0)})$.

An Alternating Optimization

- Let $\mathbf{u}^{(k)} = (\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)})$ for $k \geq 0$, defined as follows.
- Let $\mathbf{u}_1^{(0)}$ be an arbitrarily chosen vector in A_1, and let $\mathbf{u}_2^{(0)} = c_2(\mathbf{u}_1^{(0)})$.
- For $k \geq 1$, $\mathbf{u}^{(k)}$ is defined by

$$\mathbf{u}_1^{(k)} = c_1(\mathbf{u}_2^{(k-1)})$$

and

$$\mathbf{u}_2^{(k)} = c_2(\mathbf{u}_1^{(k)}).$$

An Alternating Optimization

- Let $\mathbf{u}^{(k)} = (\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)})$ for $k \geq 0$, defined as follows.
- Let $\mathbf{u}_1^{(0)}$ be an arbitrarily chosen vector in A_1, and let $\mathbf{u}_2^{(0)} = c_2(\mathbf{u}_1^{(0)})$.
- For $k \geq 1$, $\mathbf{u}^{(k)}$ is defined by

$$\mathbf{u}_1^{(k)} = c_1(\mathbf{u}_2^{(k-1)})$$

  and

$$\mathbf{u}_2^{(k)} = c_2(\mathbf{u}_1^{(k)}).$$

- Let

$$f^{(k)} = f(\mathbf{u}^{(k)}).$$

An Alternating Optimization

- Let $\mathbf{u}^{(k)} = (\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)})$ for $k \geq 0$, defined as follows.
- Let $\mathbf{u}_1^{(0)}$ be an arbitrarily chosen vector in A_1, and let $\mathbf{u}_2^{(0)} = c_2(\mathbf{u}_1^{(0)})$.
- For $k \geq 1$, $\mathbf{u}^{(k)}$ is defined by

$$\mathbf{u}_1^{(k)} = c_1(\mathbf{u}_2^{(k-1)})$$

  and

$$\mathbf{u}_2^{(k)} = c_2(\mathbf{u}_1^{(k)}).$$

- Let

$$f^{(k)} = f(\mathbf{u}^{(k)}).$$

- Then

$$f^{(k)} \geq f^{(k-1)}.$$

$$f(\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}) \qquad f(\mathbf{u}^{(0)}) = f^{(0)}$$

$$f(\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)})$$

$$\downarrow$$

$$f(\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(0)})$$

$$f(\mathbf{u}^{(0)}) = f^{(0)}$$

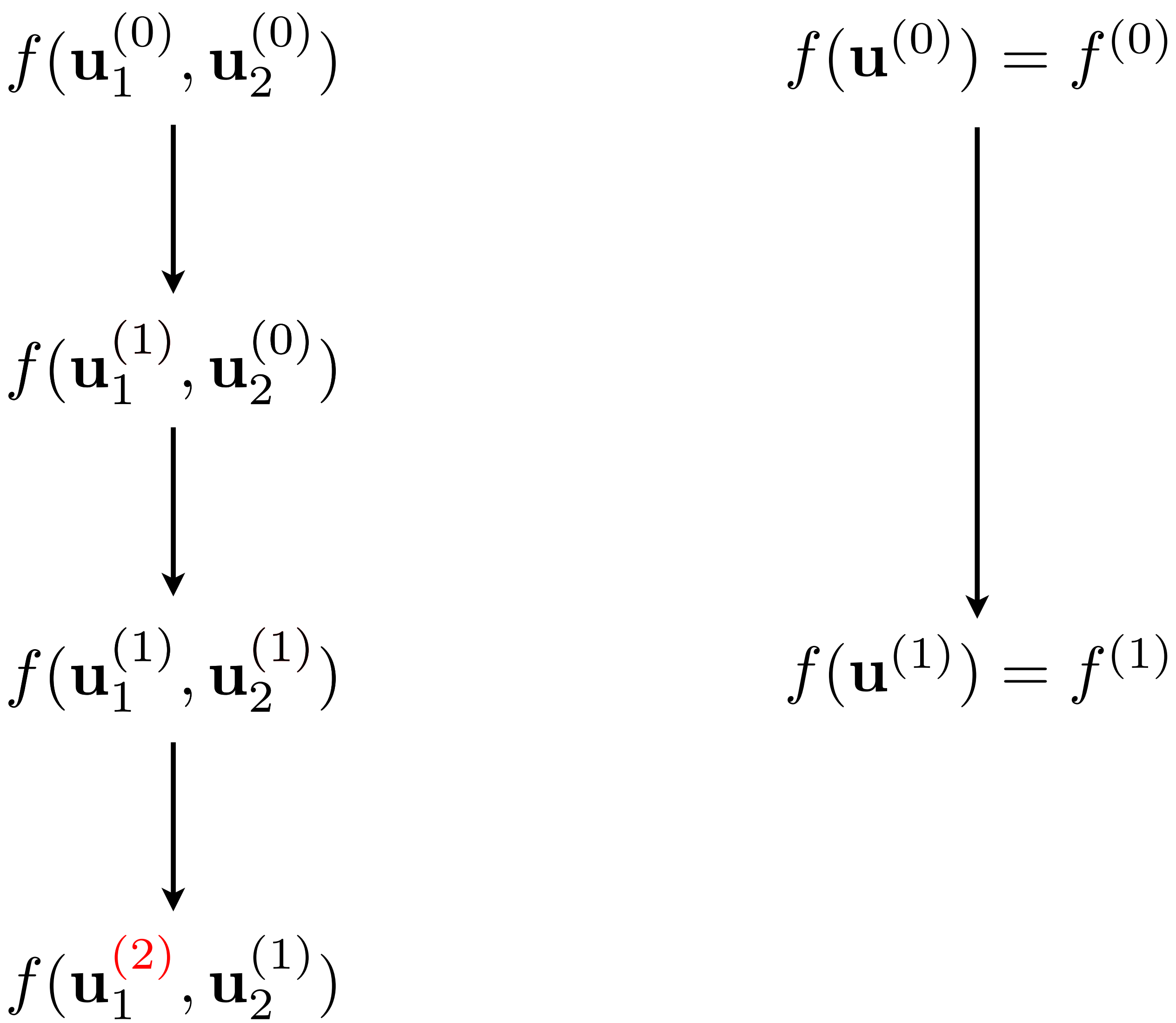
$$f(\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}) \qquad\qquad f(\mathbf{u}^{(0)}) = f^{(0)}$$

$$\downarrow$$

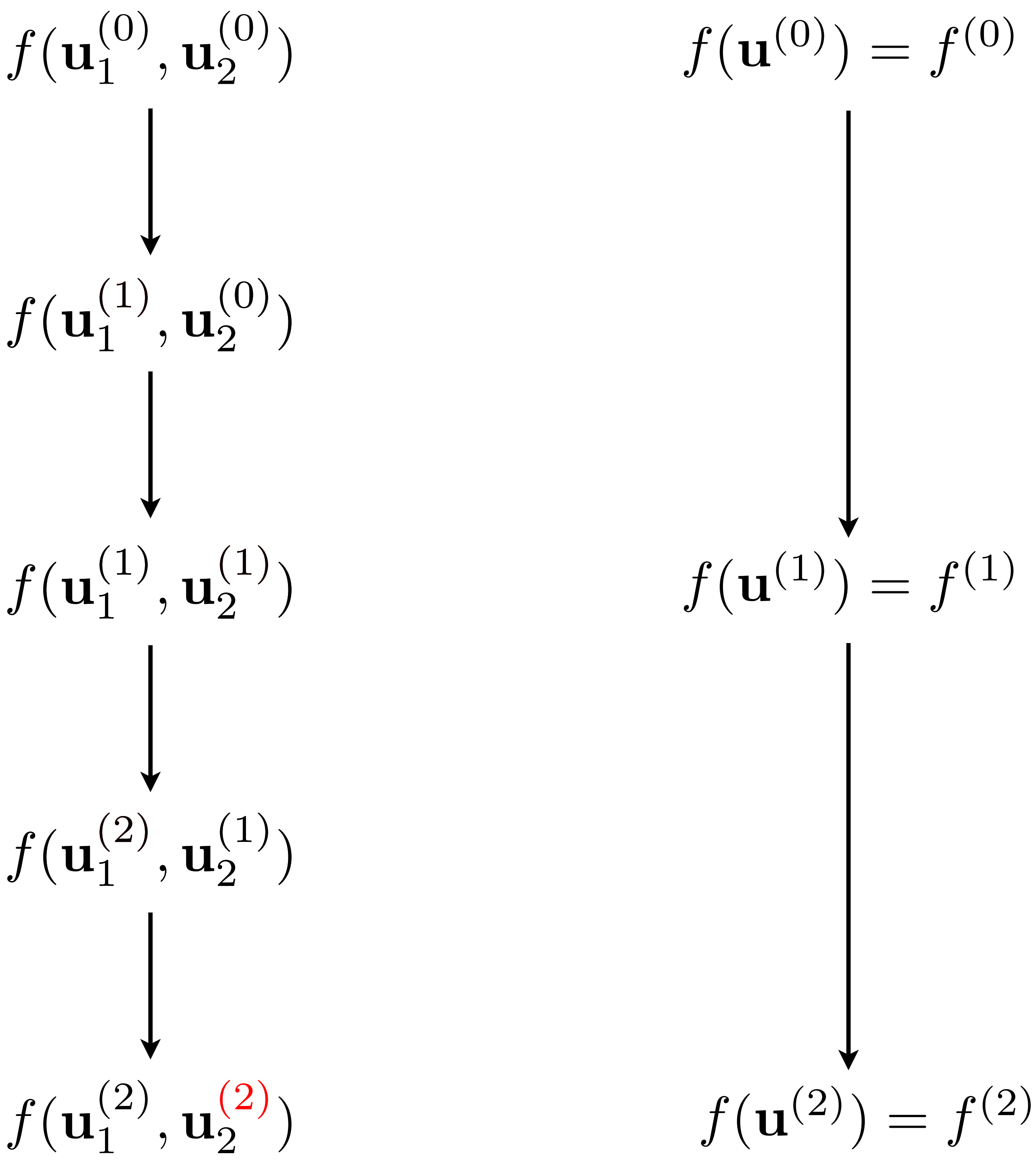
$$f(\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(0)})$$

$$\downarrow$$

$$f(\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(1)}) \qquad\qquad f(\mathbf{u}^{(1)}) = f^{(1)}$$



$f(\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)})$
$f(\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(0)})$
$f(\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(1)})$
$f(\mathbf{u}_1^{(2)}, \mathbf{u}_2^{(1)})$
$f(\mathbf{u}^{(0)}) = f^{(0)}$
$f(\mathbf{u}^{(1)}) = f^{(1)}$



$f(\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)})$
$f(\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(0)})$
$f(\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(1)})$
$f(\mathbf{u}_1^{(2)}, \mathbf{u}_2^{(1)})$
$f(\mathbf{u}_1^{(2)}, \mathbf{u}_2^{(2)})$
$f(\mathbf{u}^{(0)}) = f^{(0)}$
$f(\mathbf{u}^{(1)}) = f^{(1)}$
$f(\mathbf{u}^{(2)}) = f^{(2)}$

- Since the sequence $f^{(k)}$ is non-decreasing, it must converge because f is bounded from above.

- Since the sequence $f^{(k)}$ is non-decreasing, it must converge because f is bounded from above.
- We will show that $f^{(k)} \to f^*$ if f is concave.

- Since the sequence $f^{(k)}$ is non-decreasing, it must converge because f is bounded from above.
- We will show that $f^{(k)} \to f^*$ if f is concave.
- Replacing f by $-f$, the double supremum becomes the double infimum

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

- Since the sequence $f^{(k)}$ is non-decreasing, it must converge because f is bounded from above.

- We will show that $f^{(k)} \to f^*$ if f is concave.

- Replacing f by $-f$, the double supremum becomes the double infimum

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

- The same alternating optimization algorithm can be applied to compute this infimum.

- Since the sequence $f^{(k)}$ is non-decreasing, it must converge because f is bounded from above.

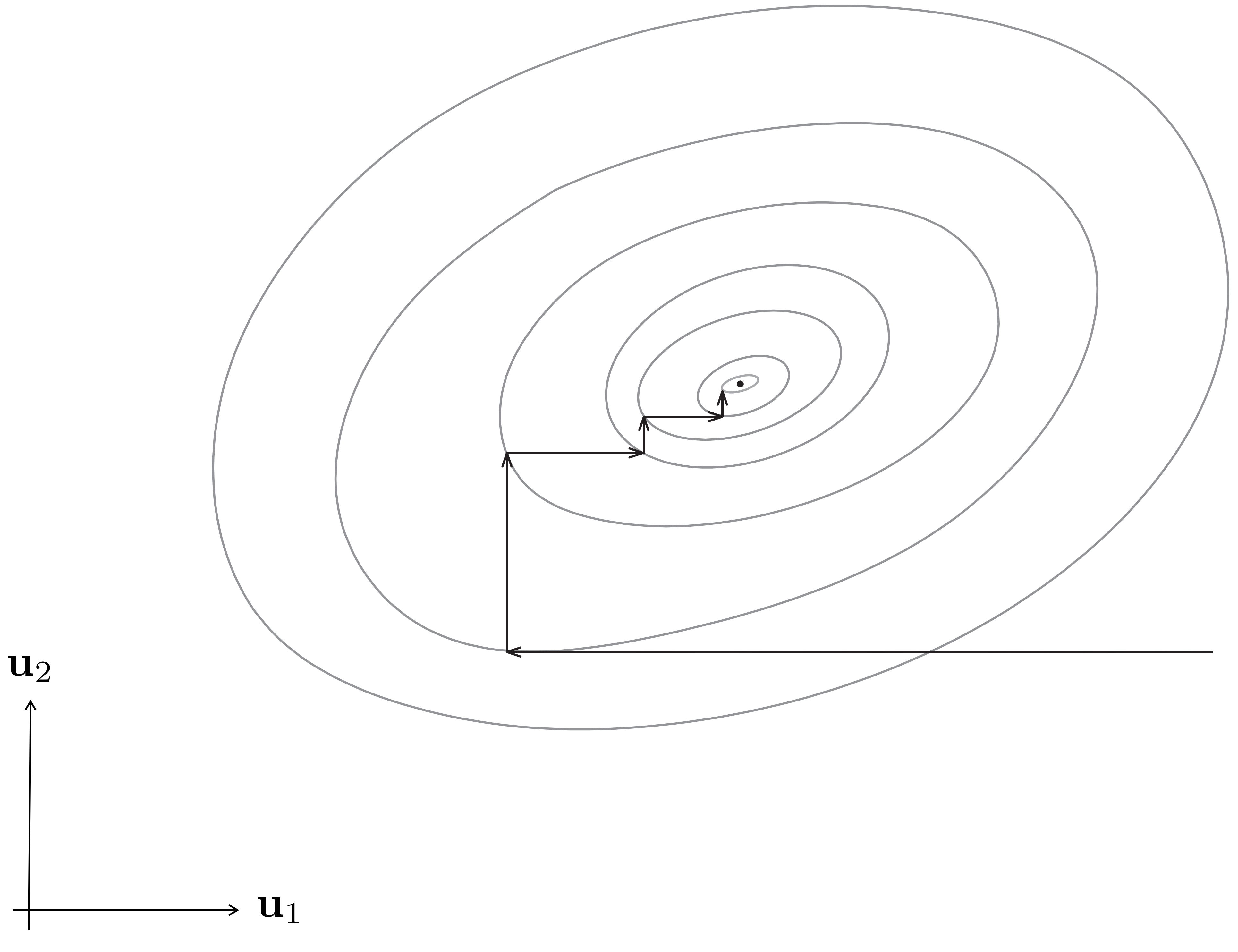
- We will show that $f^{(k)} \to f^*$ if f is concave.

- Replacing f by $-f$, the double supremum becomes the double infimum

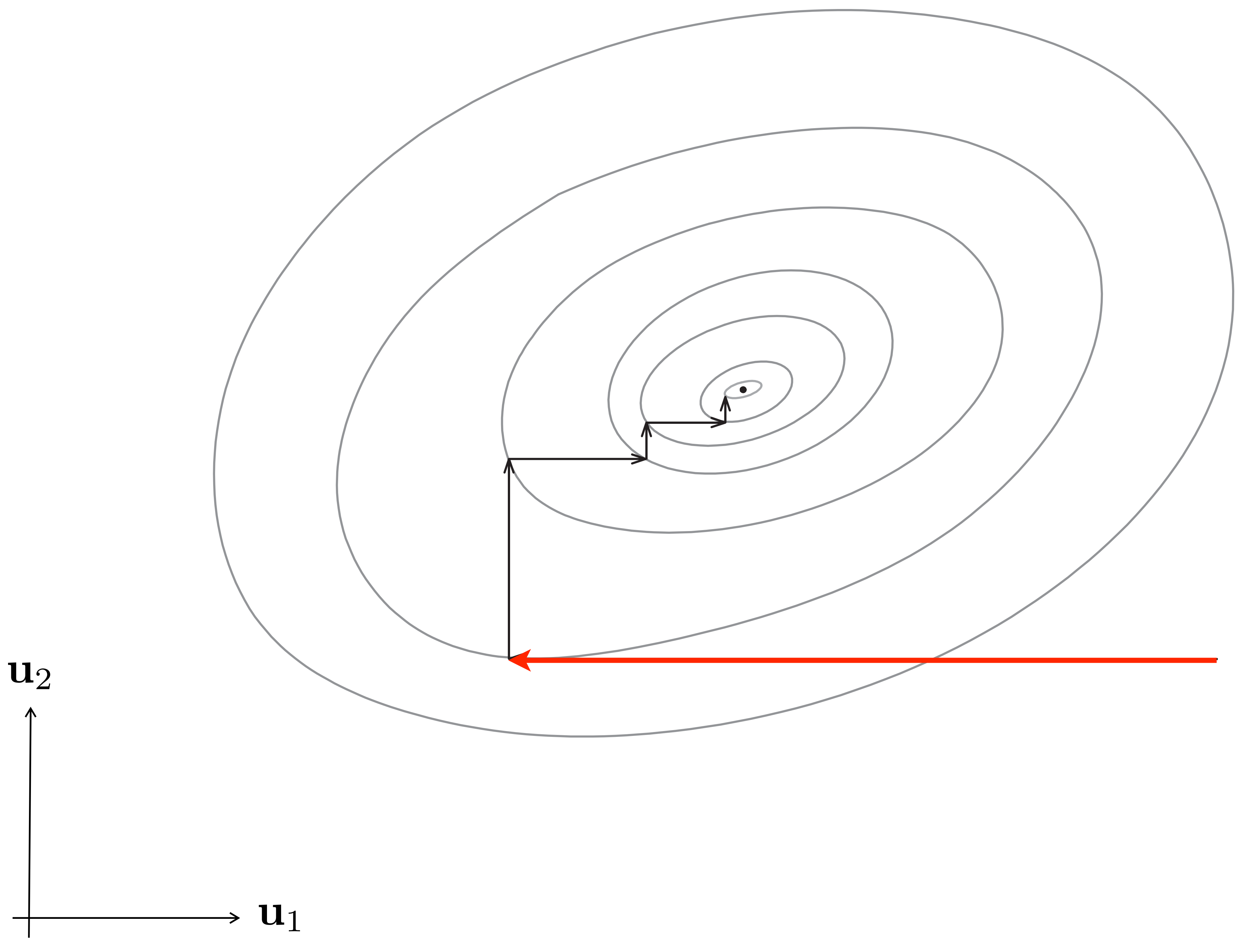
$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

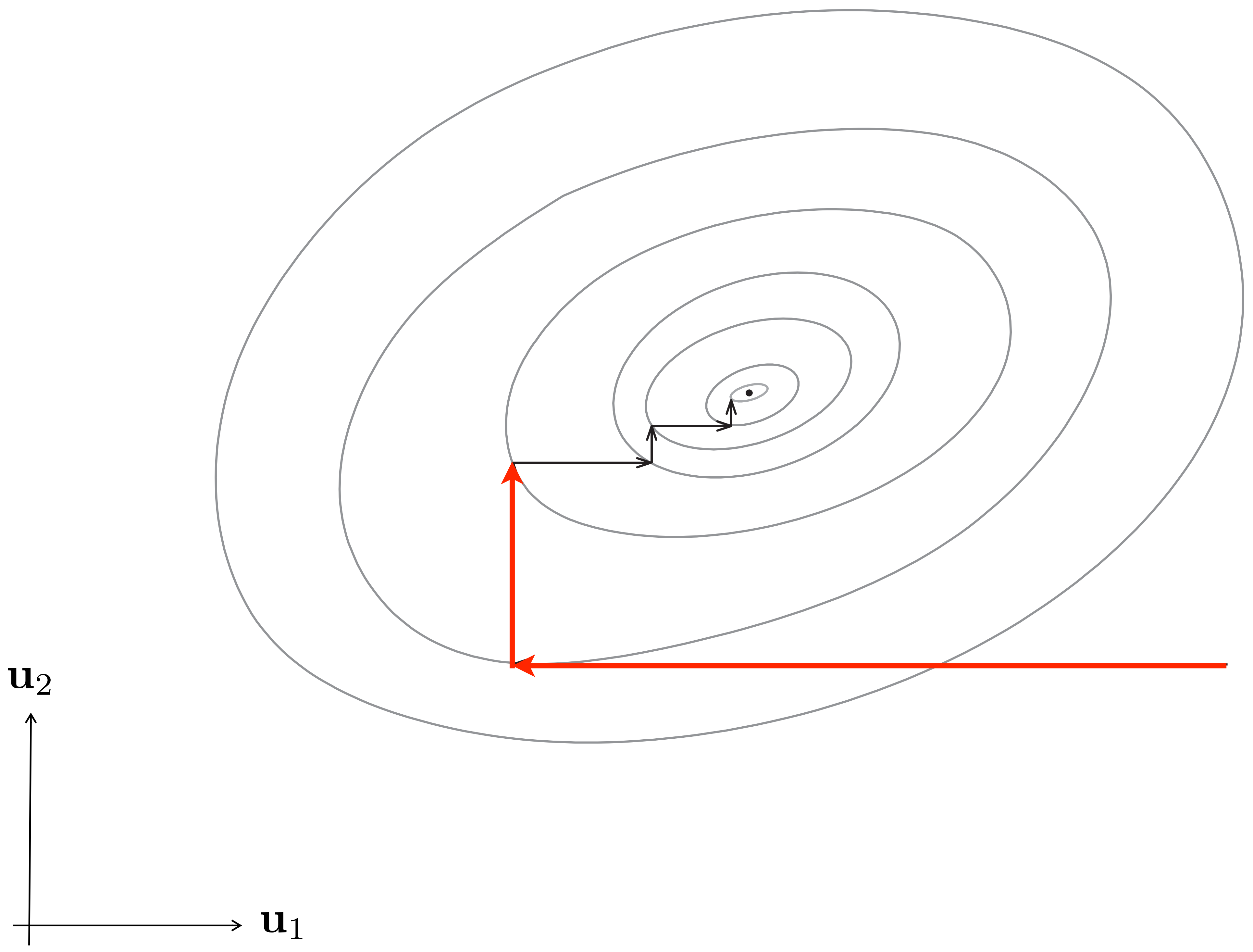
- The same alternating optimization algorithm can be applied to compute this infimum.

- The alternating optimization algorithm will be specialized for computing C and $R(D)$.

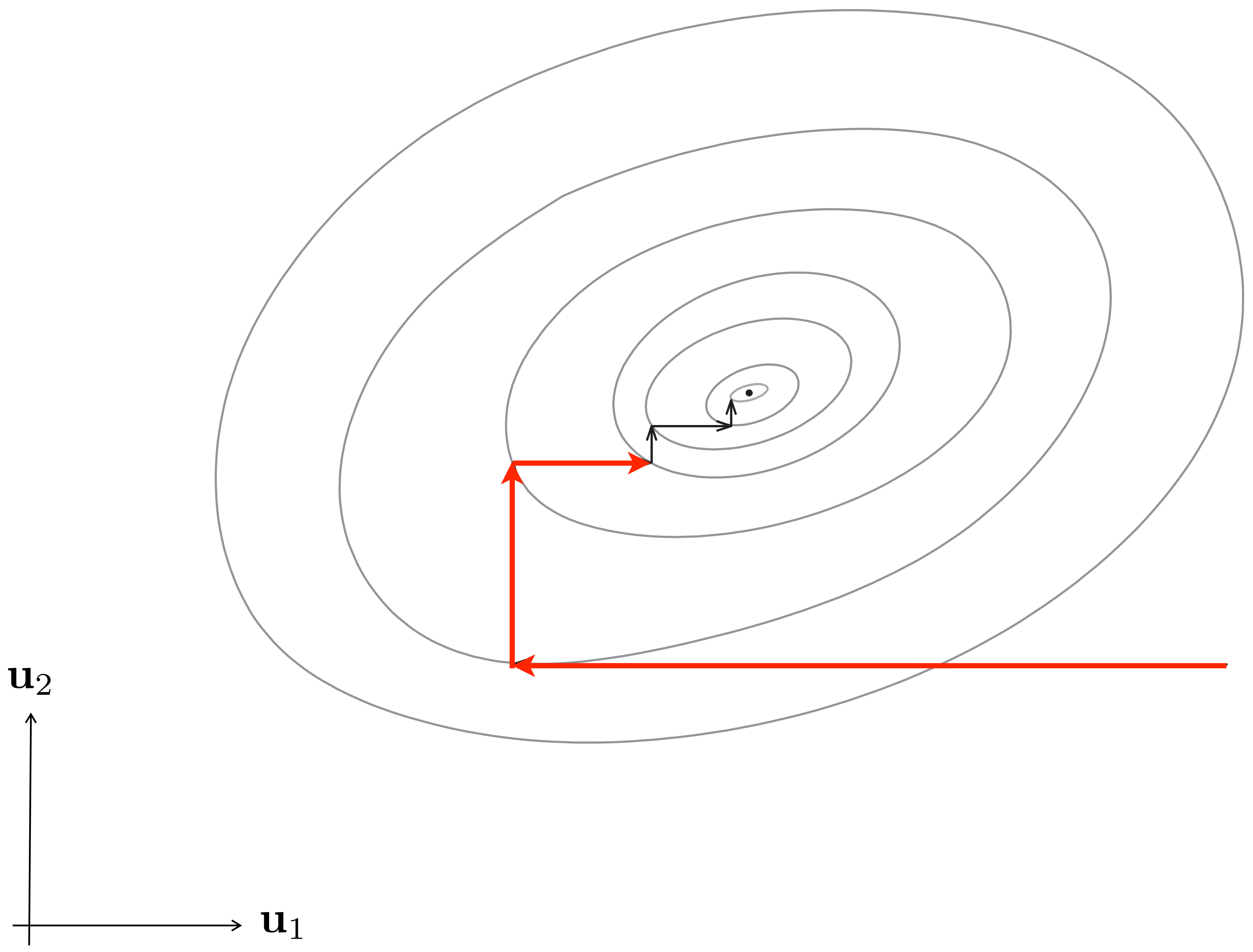


$\mathbf{u}_2$
$\mathbf{u}_1$

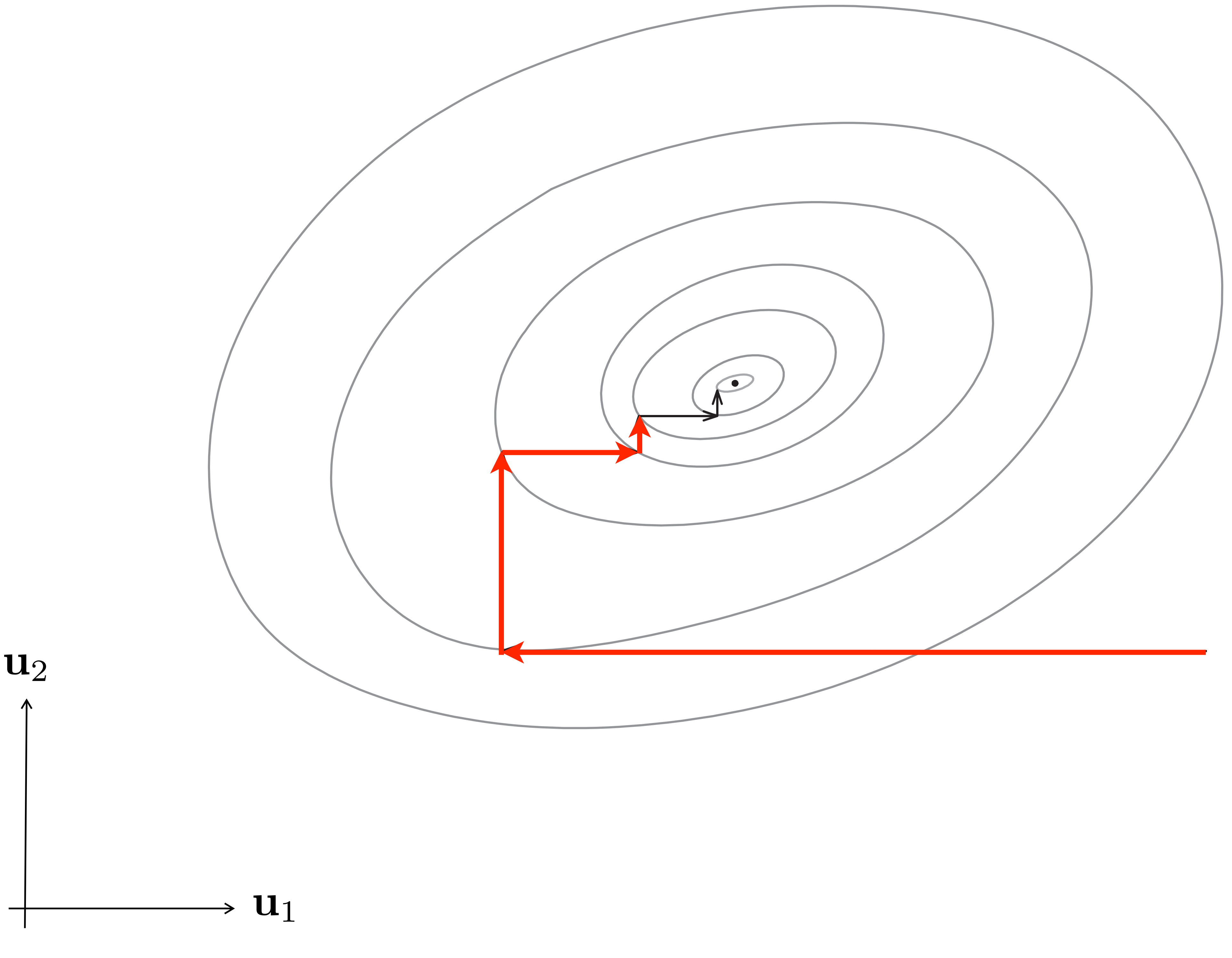

$\mathbf{u}_2$
$\mathbf{u}_1$

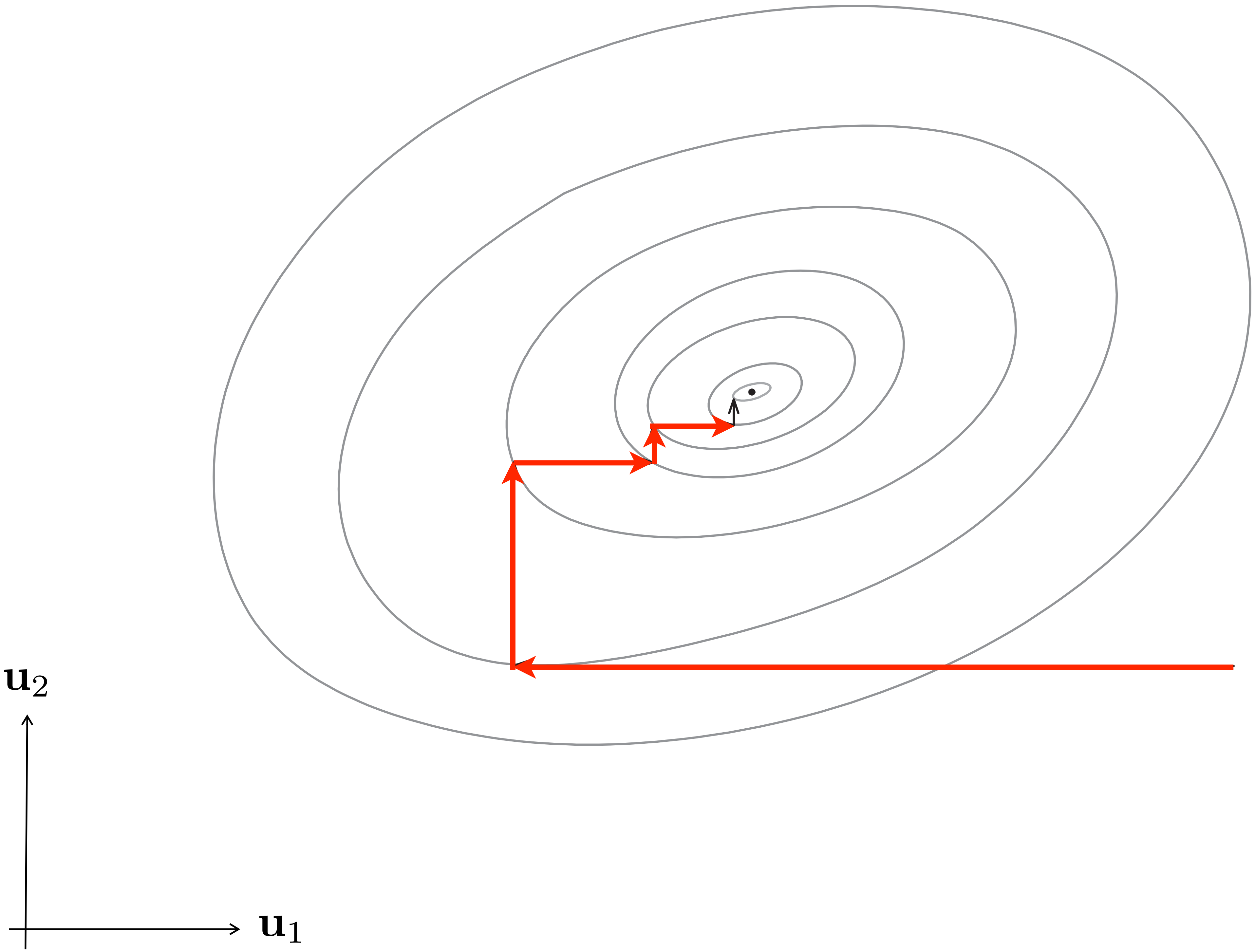


$\mathbf{u}_2$
$\mathbf{u}_1$

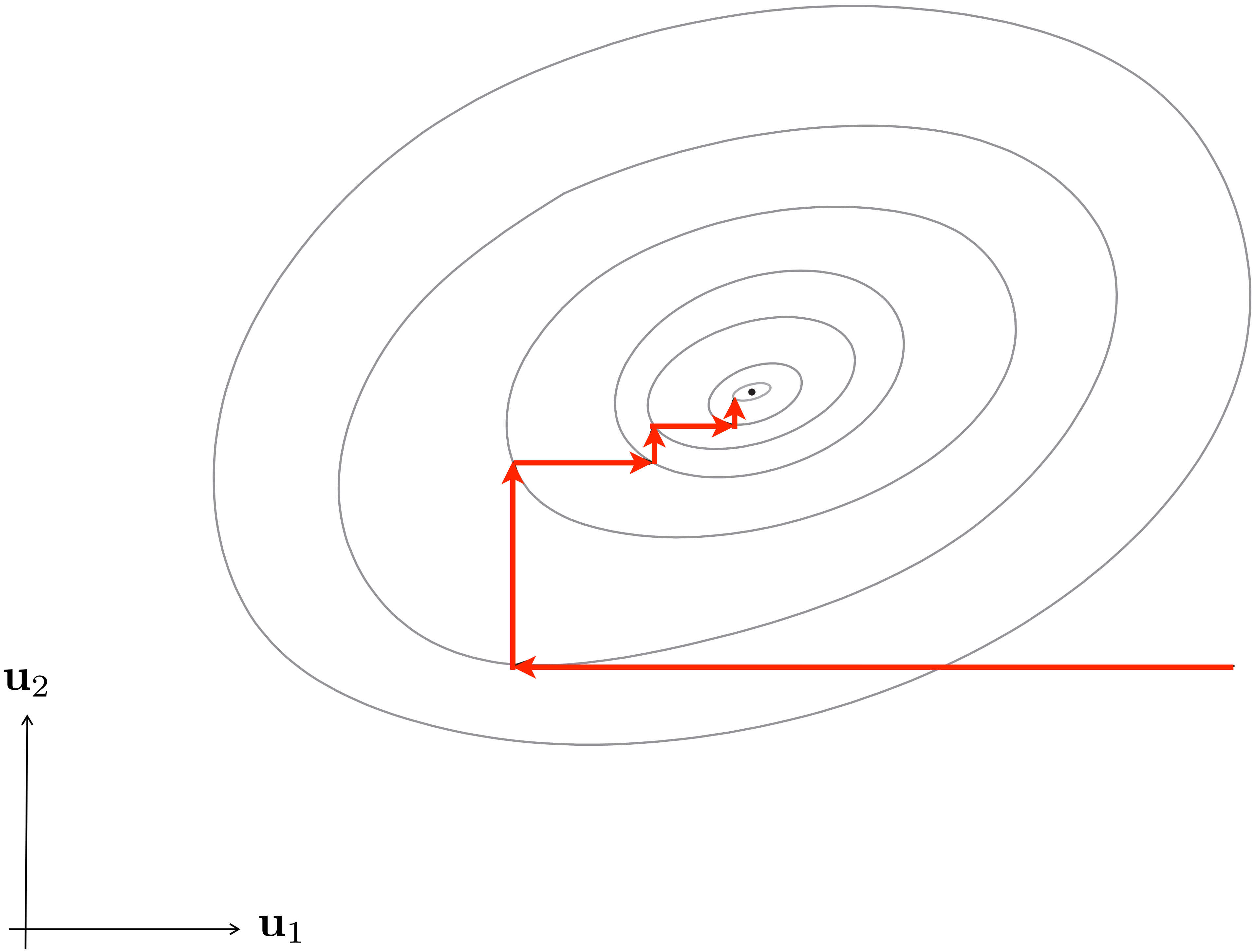


$\mathbf{u}_2$
$\mathbf{u}_1$


$\mathbf{u}_2$
$\mathbf{u}_1$



$\mathbf{u}_2$
$\mathbf{u}_1$



$\mathbf{u}_2$
$\mathbf{u}_1$