Math prelims – Linear algebra & Multivariable calculus

MATH 2740 – Mathematics of Data Science – Lecture 04

Julien Arino

julien.arino@umanitoba.ca

Department of Mathematics @ University of Manitoba

Fall 202X

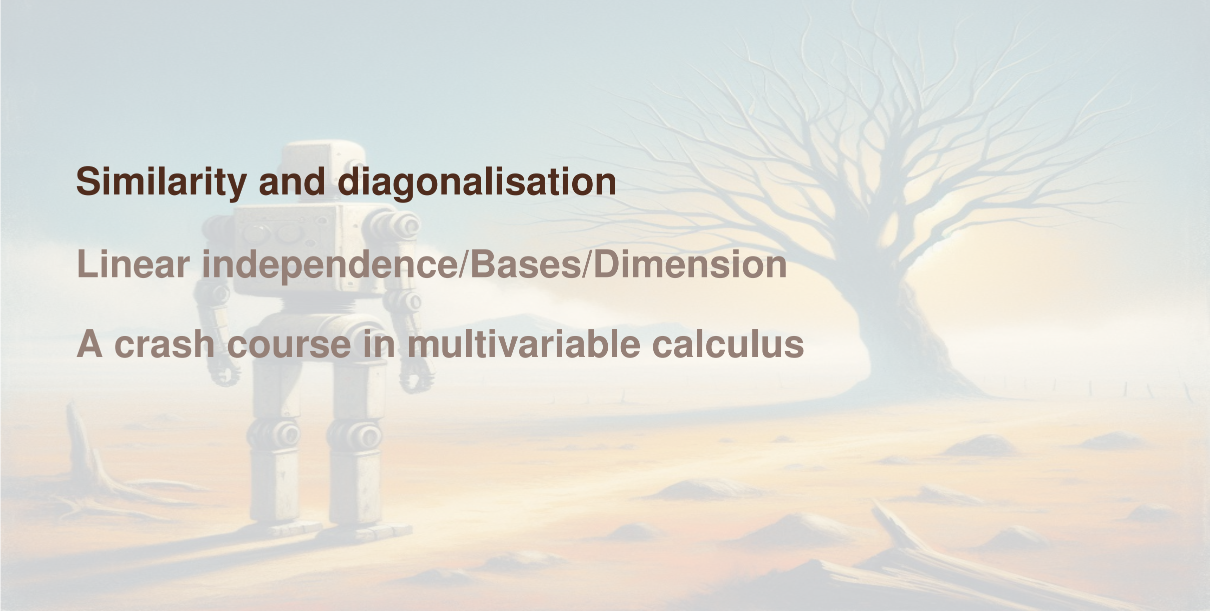
The University of Manitoba campuses are located on original lands of Anishinaabeg, Ininew, Anisininew, Dakota and Dene peoples, and on the National Homeland of the Red River Métis. We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of Reconciliation and collaboration.

Outline

Similarity and diagonalisation

Linear independence/Bases/Dimension

A crash course in multivariable calculus



Similarity and diagonalisation

Linear independence/Bases/Dimension

A crash course in multivariable calculus

Similarity

Definition 1 (Similarity)

$A, B \in \mathcal{M}_n$ are **similar** ($A \sim B$) if $\exists P \in \mathcal{M}_n$ invertible s.t.

$$P^{-1}AP = B$$

Theorem 2 ($\sim$ is an equivalence relation)

$A, B, C \in \mathcal{M}_n$, then

- $A \sim A$ ($\sim$ ***reflexive***)
- $A \sim B \implies B \sim A$ ($\sim$ ***symmetric***)
- $A \sim B$ *and* $B \sim C \implies A \sim C$ ($\sim$ ***transitive***)

Similarity (cont.)

Theorem 3

$A, B \in \mathcal{M}_n$ with $A \sim B$. Then

- $\det A = \det B$
- *A invertible $\iff$ B invertible*
- *A and B have the same eigenvalues*

Diagonalisation

Definition 4 (Diagonalisability)

$A \in \mathcal{M}_n$ is **diagonalisable** if $\exists D \in \mathcal{M}_n$ diagonal s.t. $A \sim D$

In other words, $A \in \mathcal{M}_n$ is diagonalisable if there exists a diagonal matrix $D \in \mathcal{M}_n$ and a nonsingular matrix $P \in \mathcal{M}_n$ s.t. $P^{-1}AP = D$

Could of course write $PAP^{-1} = D$ since P invertible, but $P^{-1}AP$ makes more sense for computations

Theorem 5

$A \in \mathcal{M}_n$ diagonalisable $\iff$ A has n linearly independent eigenvectors

Corollary 6 (Sufficient condition for diagonalisability)

$A \in \mathcal{M}_n$ has all its eigenvalues distinct $\implies$ A diagonalisable

For $P^{-1}AP = D$: in P, put the linearly independent eigenvectors as columns and in D, the corresponding eigenvalues


Similarity and diagonalisation
Linear independence/Bases/Dimension
A crash course in multivariable calculus

Linear combination and span

Definition 7 (Linear combination)

Let V be a vector space. A **linear combination** of a set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ of vectors in V is a *vector*

$$c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k$$

where $c_1, \ldots, c_k \in \mathbb{F}$

Definition 8 (Span)

The set of all linear combinations of a set of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is the **span** of $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$,

$$\operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \{c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k : c_1, \ldots, c_k \in \mathbb{F}\}$$

Finite/infinite-dimensional vector spaces

Theorem 9

The span of a set of vectors in V is the smallest subspace of V containing all the vectors in the set

Definition 10 (Set of vectors spanning a space)

If $\text{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = V$, we say $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ **spans** V

Definition 11 (Dimension of a vector space)

A vector space V is **finite-dimensional** if some set of vectors in it spans V. A vector space V is **infinite-dimensional** if it is not finite-dimensional

Linear (in)dependence

Definition 12 (Linear independence/Linear dependence)

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of vectors in a vector space V is **linearly independent** if

$$(c_1 \mathbf{v}_1 + \cdots + c_k \mathbf{v}_k = 0) \Leftrightarrow (c_1 = \cdots = c_k = 0)\,,$$

where $c_1, \ldots, c_k \in \mathbb{F}$. A set of vectors is **linearly dependent** if it is not linearly independent.

If linearly dependent, assume w.l.o.g. that $c_1 \neq 0$, then

$$\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_k}{c_1}\mathbf{v}_k$$

i.e., $\mathbf{v}_1$ is a linear combination of the other vectors in the set

Theorem 13

Let V be a finite-dimensional vector space. Then the ***cardinal*** *(number of elements) of every linearly independent set of vectors is less than or equal to the number of elements in every spanning set of vectors*

E.g., in $\mathbb{R}^3$, a set with 4 or more vectors is automatically linearly dependent

Basis

Definition 14 (Basis)

Let V be a vector space. A **basis** of V is a set of vectors in V that is both linearly independent and spanning

Theorem 15 (Criterion for a basis)

A set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ of vectors in a vector space V is a basis of V $\iff$ $\forall \boldsymbol{v} \in V$, $\boldsymbol{v}$ can be written uniquely in the form

$$\boldsymbol{v} = c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k,$$

where $c_1, \dots, c_k \in \mathbb{F}$

Plus/Minus Theorem

Theorem 16 (Plus/Minus Theorem)

S a nonempty set of vectors in vector space V

- *If S is linearly independent and $V \ni \boldsymbol{v} \notin \operatorname{span}(S)$, then $S \cup \{\boldsymbol{v}\}$ is linearly independent*
- *If $\boldsymbol{v} \in S$ is linear combination of other vectors in S, then* $\operatorname{span}(S) = \operatorname{span}(S - \{\boldsymbol{v}\})$

More on bases

Theorem 17 (Basis of finite-dimensional vector space)

Every finite-dimensional vector space has a basis

Theorem 18

Any two bases of a finite-dimensional vector space have the same number of vectors

Definition 19 (Dimension)

The **dimension** $\dim V$ of a finite-dimensional vector space V is the number of vectors in any basis of the vector space

Theorem 20 (Dimension of a subspace)

Let V be a finite-dimensional vector space and $U \subset V$ be a subspace of V. Then $\dim U \leq \dim V$

Constructing bases

Theorem 21

Let V be a finite-dimensional vector space. Then every linearly independent set of vectors in V with $\dim V$ elements is a basis of V

Theorem 22

Let V be a finite-dimensional vector space. Then every spanning set of vectors in V with $\dim V$ elements is a basis of V

Linear independence/Bases/Dimension

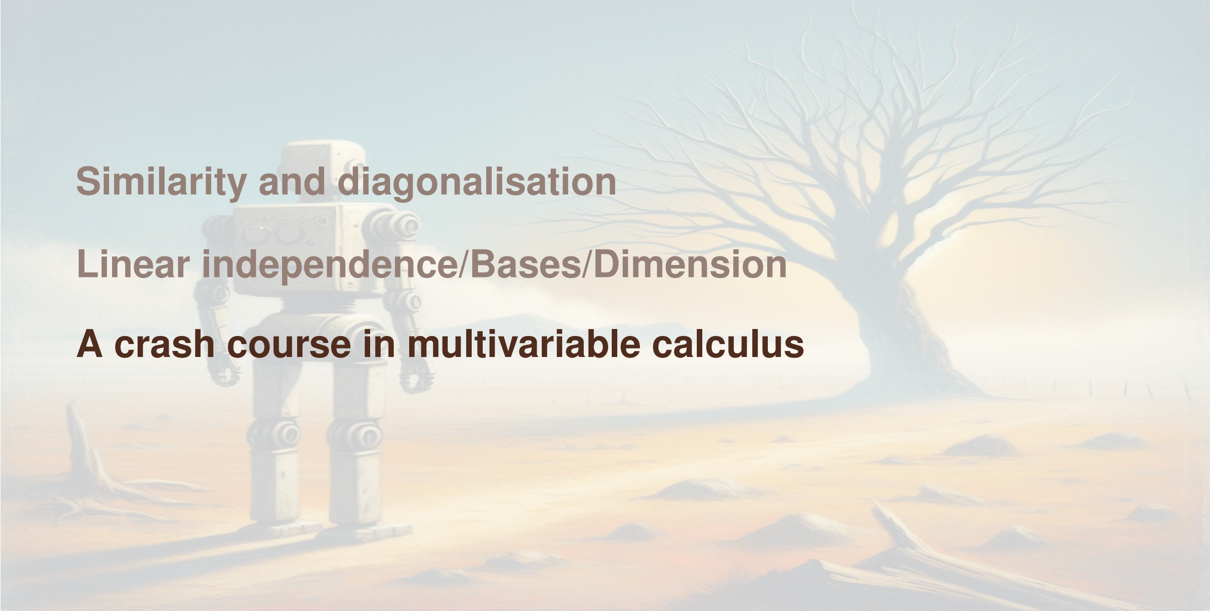
Linear algebra in a nutshell

To finish: the “famous” “growing result”

Theorem 23

Let $A \in \mathcal{M}_n$. The following statements are equivalent (TFAE)

1. *The matrix A is invertible*
2. $\forall \boldsymbol{b} \in \mathbb{F}^n$, $A\boldsymbol{x} = \boldsymbol{b}$ *has a unique solution* ($\boldsymbol{x} = A^{-1}\boldsymbol{b}$)
3. *The only solution to* $A\boldsymbol{x} = \mathbf{0}$ *is the trivial solution* $\boldsymbol{x} = \mathbf{0}$
4. $RREF(A) = \mathbb{I}_n$
5. *The matrix A is equal to a product of elementary matrices*
6. $\forall \boldsymbol{b} \in \mathbb{F}^n$, $A\boldsymbol{x} = \boldsymbol{b}$ *has a solution*
7. *There is a matrix $B \in \mathcal{M}_n$ such that $AB = \mathbb{I}_n$*
8. *There is an invertible matrix $B \in \mathcal{M}_n$ such that $AB = \mathbb{I}_n$*
9. $\det(A) \neq 0$
10. 0 *is not an eigenvalue of A*



Similarity and diagonalisation

Linear independence/Bases/Dimension

A crash course in multivariable calculus

A crash course in multivariable calculus

What is Multivariable Calculus?

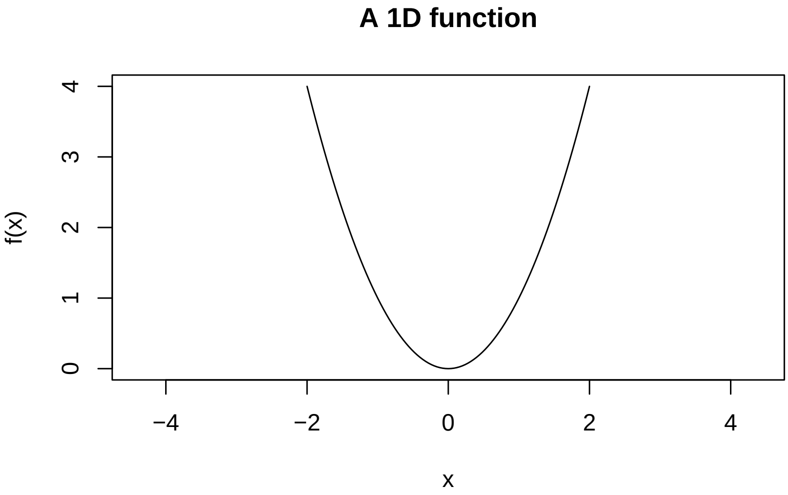
Optimization with Constraints

The Lagrangian Method

Conclusion

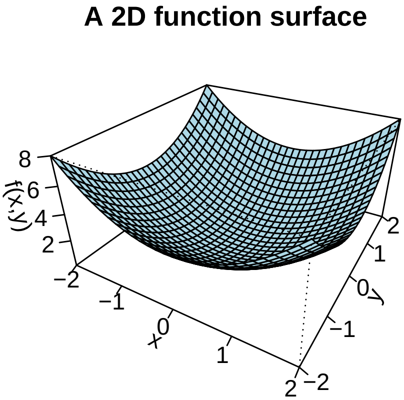
One dimension

MATH 1500 & 1700 deal with functions of one variable, like $f(x) = x^2$



Multivariable calculus

Multivariable calculus extends this to functions of two or more variables, like $f(x, y) = x^2 + y^2$



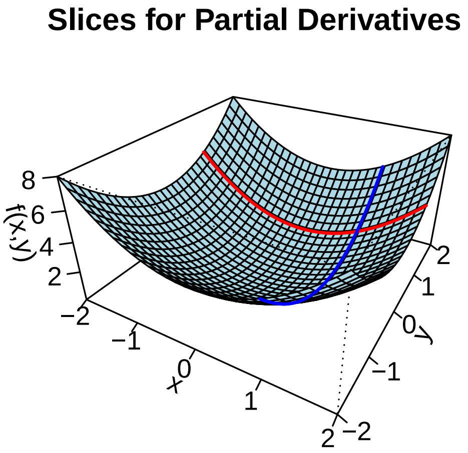
Partial derivatives

How do we measure the “slope” on a 3D surface?

A **partial derivative** measures the slope in a direction parallel to one of the axes

- $\frac{\partial f}{\partial x}$ measures height change as we move only in the x direction. Treat y as a constant
- $\frac{\partial f}{\partial y}$ measures height change as we move only in the y direction. Treat x as a constant

Partial derivatives



The Steepest path: the gradient

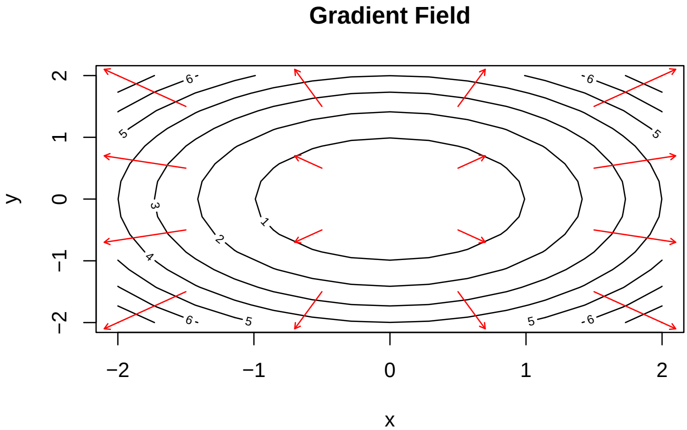
The **gradient**, denoted ∇f, is a vector that combines all the partial derivatives:

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

What does it tell us?

- **Direction:** it points in the direction of the *steepest ascent*
- **Magnitude:** its length represents the steepness of that ascent

Follow the gradient



At a peak or a valley (a local max/min), the ground is flat. So, $\nabla f = (0, 0)$

A crash course in multivariable calculus

What is Multivariable Calculus?

Optimization with Constraints

The Lagrangian Method

Conclusion

The real-world problem

Often, we want to maximize or minimize a function, but we don't have unlimited freedom. We have **constraints**

- Maximize the profit of your company... *subject to a limited budget*
- Minimize the material used for a can... *that must hold a specific volume*
- Find the highest point on a mountain... *while staying on a specific trail*

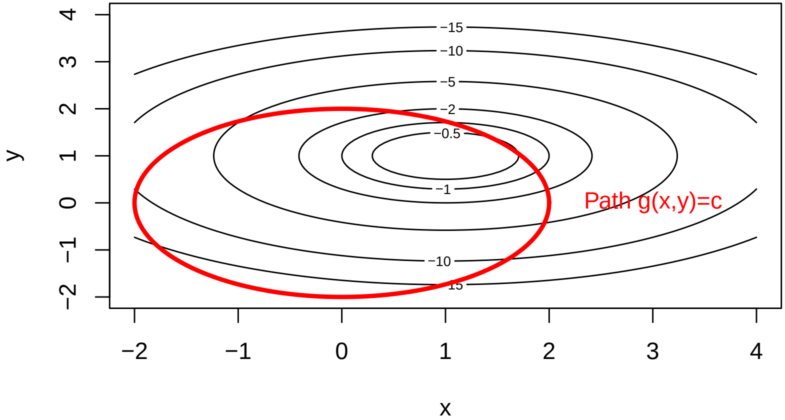
Setting the gradient to zero ($\nabla f = 0$) finds the highest point on the whole mountain, which might not be on our trail!

Visualizing the problem

Imagine our function $f(x, y)$ is the altitude on a map (contour lines)

Our constraint, $g(x, y) = c$, is a specific path we must walk on

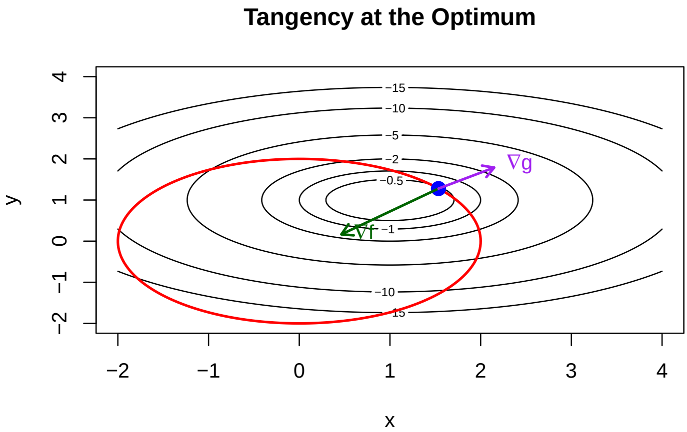
Optimization with a Constraint



We are looking for the highest (or lowest) point *along the red path*

The key insight

At the optimal point on the path, the path will be perfectly **tangent** to the contour line of the surface

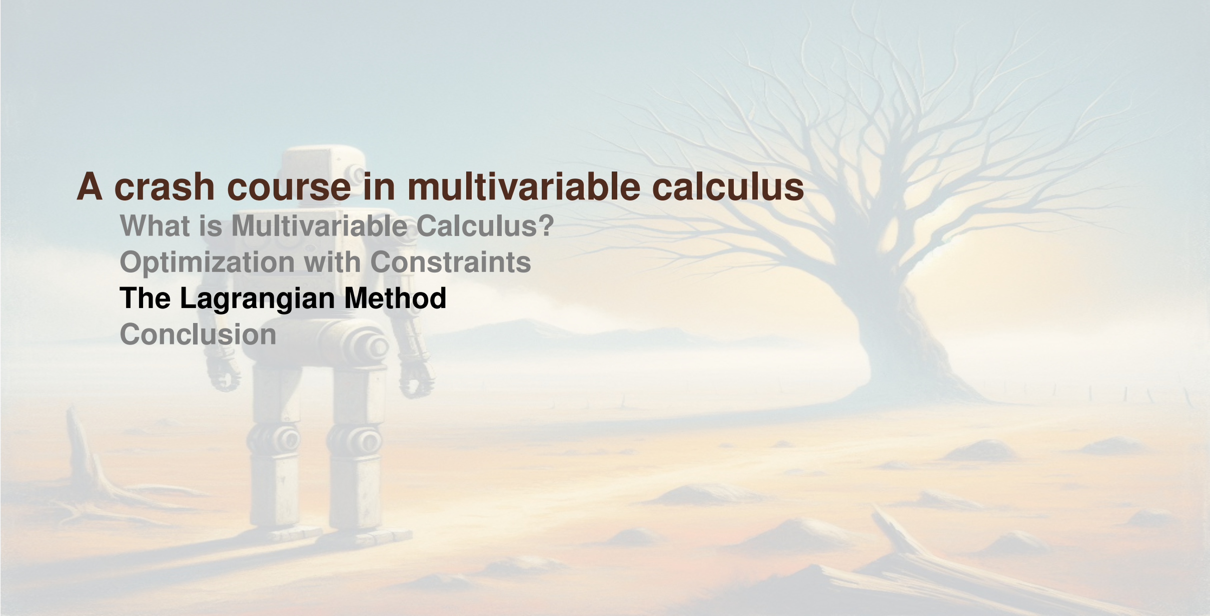


Why? If the path crossed the contour line, you could move along the path to get to a higher (or lower) contour

Mathematically, this tangency means the gradient vectors of the function and the constraint are **parallel**

$$\nabla f = \lambda \nabla g$$

The scalar λ (lambda) is called the **Lagrange multiplier**



A crash course in multivariable calculus

- What is Multivariable Calculus?
- Optimization with Constraints
- **The Lagrangian Method**
- Conclusion

The Lagrangian function

The condition $\nabla f = \lambda \nabla g$ is clever, but solving it can be messy

Instead, we combine our function and constraint into a single, new function called the **Lagrangian**

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c]$$

- $f(x, y)$ the function we want to optimize
- $g(x, y) = c$ the constraint we must follow
- λ the Lagrange multiplier

Finding the unconstrained optimum of $\mathcal{L}$ solves the original constrained problem!

The method – step-by-step

To find the optimum of the Lagrangian $\mathcal{L}(x, y, \lambda)$, we find where its gradient is zero

We take the partial derivative with respect to *all* its variables (x, y, and λ) and set them to zero

1. $\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0$
2. $\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0$
3. $\frac{\partial \mathcal{L}}{\partial \lambda} = -(g(x, y) - c) = 0 \implies g(x, y) = c$

The first two equations rearrange to $\nabla f = \lambda \nabla g$ and the third equation is the original constraint

Example: Fencing a Field

Problem: You have 40 meters of fence. What is the largest rectangular area you can enclose?

- **Maximize Area:** $A(x, y) = xy$
- **Constraint (Perimeter):** $2x + 2y = 40$

1. **Form the Lagrangian:**

$$\mathcal{L}(x, y, \lambda) = xy - \lambda(2x + 2y - 40)$$

2. **Take Partial Derivatives:**

- $\frac{\partial \mathcal{L}}{\partial x} = y - 2\lambda = 0 \implies y = 2\lambda$
- $\frac{\partial \mathcal{L}}{\partial y} = x - 2\lambda = 0 \implies x = 2\lambda$
- $\frac{\partial \mathcal{L}}{\partial \lambda} = -(2x + 2y - 40) = 0$

Example: Solution

From the first two equations, we see that $x = y$

Now, substitute this into the third equation (the constraint):

$$2x + 2(x) = 40$$

$$4x = 40$$

$$x = 10$$

Since $x = y$, we have $y = 10$, i.e., optimal dimensions are 10m by 10m (a square), giving a maximum area of 100 m^2

A crash course in multivariable calculus

What does λ mean?

The Lagrange multiplier λ has a very useful interpretation

It tells you how much the optimal value of your function f will change if you slightly relax the constraint c

$$\lambda = \frac{df_{optimal}}{dc}$$

In our example: If we had 41 meters of fence instead of 40 (so c changes by 1), how much would the max area increase?

$x = y = 2\lambda$, so $\lambda = x/2 = 10/2 = 5$. The maximum area would increase by approximately 5 m^2