Lagrange Multipliers

Optimization
with Constraints

"As long as algebra and geometry have been separated, their progress have been slow and their uses limited;
but when these two sciences have been united, they have lent each mutual forces, and have marched together towards perfection."

Motivation

- Many problems in bioinformatics require us to find the maximum or minimum of a differentiable function f.
  - For example, in this course we want to find the optimal rotation that causes one protein to be superimposed over another in 3D space.
    - We will also need to do a type of optimization when we look at classification problems.

Introduction (1)

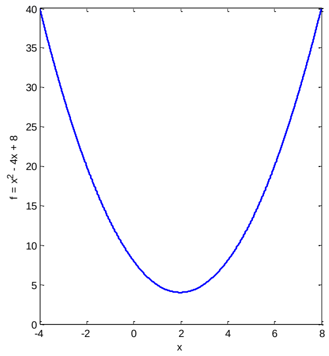
- The points in the domain of f where the minimum or maximum occurs are called the critical points (also the extreme points).
- You have already seen examples of critical points in elementary calculus:
  - Given $f(x)$ with f differentiable, find values of x such that $f(x)$ is a local minimum (or maximum).
  - Usual approach:
    - We assume these x values are such that $f'(x) = 0$.
    - That is the derivative is zero at this critical point in the x domain.

Introduction (2)

- Our optimization problems will be more complicated for two reasons:
  1. The function f depends on several variables, that is, $f(x_1, x_2, \ldots, x_n)$.
  2. In many applications, we will have a second equation $g(x_1, x_2, \ldots, x_n) = 0$ that also must be satisfied.
     - We still want a point in $\mathbb{R}^n$ that is an extreme point for f but it must also lie on the line or surface defined by g.
     - We will consider three optimization problems corresponding to $n = 1, 2, 3$ with no constraint and with a constraint (6 examples altogether).

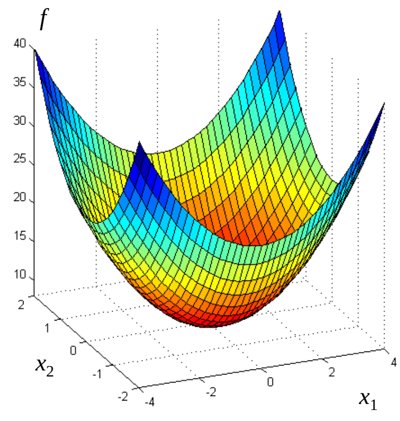
Three Optimization Problems (1)

- Problem 1 ($n = 1$):
  - Find x that will minimize $f(x) = x^2 - 4x + 8.$
    - Since $$f(x) = (x-2)^2 + 4.$$ it is clear that the minimum is at $x = 2$ and the f value is 4.
    - Note also that $$f'(x) = 2x - 4$$ and setting this to zero also gives us $x = 2$.



Three Optimization Problems (2)

- Problem 2 ($n = 2$):
  - Minimize $f(x) = x_1^2 + 4x_2^2 + 8.$
    - Because the squared terms cannot be negative, the minimum occurs when they are both zero, i.e. when $x_1 = 0, x_2 = 0$ and at this critical point the f value is 8.
    - Note also that $$\frac{\partial f}{\partial x_1} = 2x_1 \text{ and } \frac{\partial f}{\partial x_2} = 8x_2$$ Setting both of these to zero also gives us $x_1 = 0, x_2 = 0$.



Three Optimization Problems (3)

- Problem 3 (n = 3):
  - Minimize $f(x) = x_1^2 + 4x_2^2 + 16x_3^2 - 8x_2 + 24.$
    - We can rewrite this as $f(x) = x_1^2 + 4(x_2 - 1)^2 + 16x_3^2 + 20$
    - Because the squared terms cannot be negative, the minimum will occur when they are all zero, that is, when $x_1 = 0, x_2 = 1, x_3 = 0$ and at this critical point the f value is 20.
    - Note also that

      $$\frac{\partial f}{\partial x_1} = 2x_1, \quad \frac{\partial f}{\partial x_2} = 8x_2 - 8 \quad \text{and} \quad \frac{\partial f}{\partial x_3} = 32x_3$$

      Setting all of these to zero also gives us $x_1 = 0, x_2 = 1, x_3 = 0$.
    - Unfortunately, a graph of f versus x_1, x_2, x_3 would be in a four dimensional space and impossible to visualize.

Optimization Via Derivatives

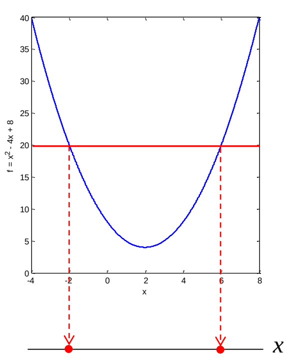
- For these simple cases (no constraint function) the general strategy is to compute the derivatives with respect to all the independent variables: $\frac{\partial f}{\partial x_i} \quad i = 1, 2, \ldots, n$
- Setting these to zero: $\frac{\partial f}{\partial x_i} = 0 \quad i = 1, 2, \ldots, n$ gives n equations in n unknowns.
- Solve the equations to get the critical points.

Level Sets (1)

- Level sets provide a visual aid for some of the geometric arguments that we will need to make.
- Given some particular real value, for example, r, a level set is a set of points in the domain of f such that f is equal to this value r.
  - Formally: $\{(x_1, x_2, \ldots, x_n) \mid f(x_1, x_2, \ldots, x_n) = r\}$.
    - For Problem 1, the level set consists of two points when $r > 4$ and a single point when $r = 4$ (empty otherwise).
    - For Problem 2, the level set is an ellipse when $r > 8$.
    - For Problem 3, the level set is an ellipsoid when $r > 20$.

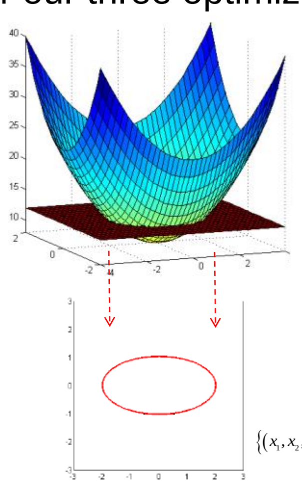
Level Sets (2)

- The level sets for our three optimization Problems:



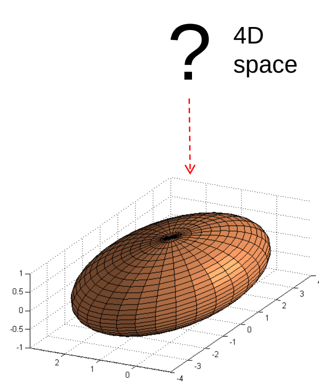
$\{x \mid f(x) = x^2 - 4x + 8 = 20\}$

$r = 20$



$\{(x_1, x_2) \mid f(x_1, x_2) = x_1^2 + 4x_2^2 + 8 = 12\}$

$r = 12$



$\{(x_1, x_2, x_3) \mid f(x_1, x_2, x_3) = x_1^2 + 4x_2^2 + 16x_3^2 - 8x_2 + 20 = 36\}$

$r = 36$

Gradient

- Given a scalar function f mapping a vector $x=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$ to a real value, that is $f(x)\in\mathbb{R}$ we define the gradient of f or grad(f) as:

$$\nabla f=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\ldots,\frac{\partial f}{\partial x_n}\right)^{\mathrm{T}}.$$

Alternate notation for $n = 3$:

$$\nabla f=\frac{\partial f}{\partial x_1}i+\frac{\partial f}{\partial x_2}j+\frac{\partial f}{\partial x_3}k.$$

  - Note that it is a column vector.
    - But its entries are functions of $x=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$.
    - We really have a vector “field”: one vector defined for *each point* in the $\mathbb{R}^n$ space.
  - SO: to get a critical point we find those $x=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$ such that $\nabla f=0$.

Gradients and Level Sets (1)

- A gradient essentially tells us how level sets change as r increases. They point in the direction of increasing r.
  - Consider a gradient vector at $x=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$ where x is a level set point.

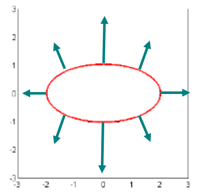
<u>Problem 1</u>

$$\nabla f=\frac{df}{dx}=2x-4$$

$x=-2\Rightarrow\frac{df}{dx}=-8$ $\quad$ $x=6\Rightarrow\frac{df}{dx}=8$

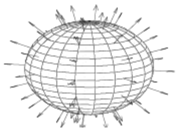
<u>Problem 2</u>

$$\nabla f=\begin{bmatrix}2x_1\\8x_2\end{bmatrix}$$



<u>Problem 3</u>

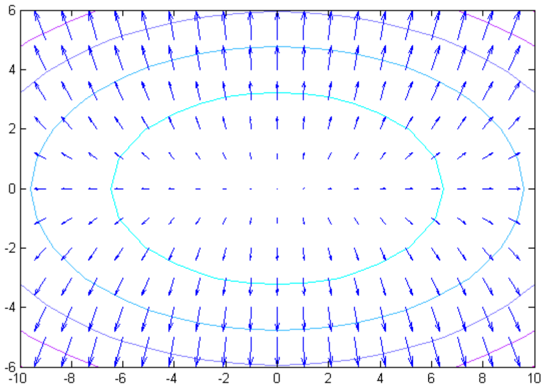
$$\nabla f=\begin{bmatrix}2x_1\\8x_2-8\\32x_3\end{bmatrix}$$



Gradients and Level Sets (2)

- In some cases a diagram will show a set of gradient vectors taken at regular intervals from the background field along with a set of level curves:

Problem 2
with an array of
gradient vectors
and 4 level curves:



Gradients and Level Sets (3)

- THEOREM: **The gradient of f is normal to the level set of f at that point.**
  - Proof sketch:
    - Consider a point p in the level set and a gradient vector that is defined at p.
    - The level set going through p is $\{x \mid f(x) = f(p)\}$.
    - Assume C is **any** differentiable curve that:
      1. Is parameterized by t $\Rightarrow C(t) = (c_1(t), c_2(t), \ldots, c_n(t))^T$.
      2. Lies within the level set of f $\Rightarrow f(c_1(t), c_2(t), \ldots, c_n(t)) = f(p)$.
      3. Passes through p when $t = t_p$ $\Rightarrow C(t_p) = (c_1(t_p), c_2(t_p), \ldots, c_n(t_p))^T = p$.

Gradients and Level Sets (4)

- THEOREM: The gradient of f is normal to the level set of f at that point.
  - Proof sketch continued:
    - The second assumption gives us

$$\frac{d}{dt}\left[f\left(c_1(t), c_2(t), \ldots, c_n(t)\right)\right] = \frac{d}{dt} f(p) = 0.$$

Since $f(p)$ is a constant.

    - But using the chain rule on the LHS we can write:

$$\sum_{i=1}^{n} \frac{\partial f\left(c_1(t), c_2(t), \ldots, c_n(t)\right)}{\partial x_i} \frac{dc_i(t)}{dt} = 0.$$

    - At $t = t_p$ this can be written as:

$$(\nabla f)^{\mathrm{T}} \left(\frac{dC}{dt}\right)\Bigg|_{t=t_p} = 0.$$

Gradient vector at p.

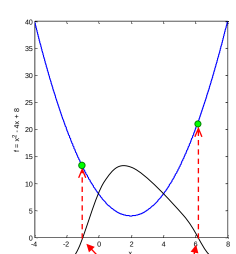
Tangent to C at p.

So, the gradient at p is normal to any tangent at p in the level set implying the gradient is normal to the level set.

Problems with Constraints (1)

- We now consider the same problems but with constraints:
  - Recall: We still want to minimize or maximize f but now the point must also satisfy the equation $g(x_1, x_2, \ldots, x_n) = 0$.
  - Let us go through the various problems and consider the effects of a constraint:
    - Problem 1:
      - Minimize $f(x) = x^2 - 4x + 8.$ subject to $g(x) = 0$.
      - Find x values such that $g(x) = 0$ then find which of these x values produces the minimal f.

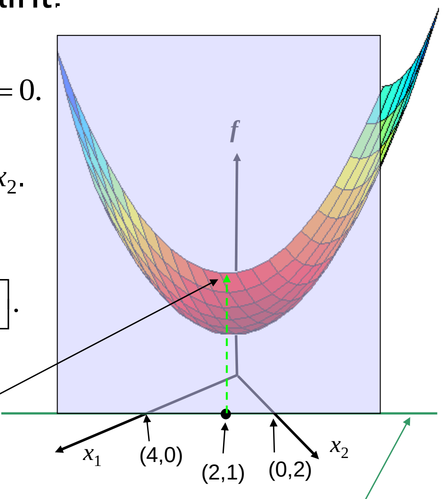
A discrete set; not very interesting.



$g(x) = 0$

Problems with Constraints (2)

- Problem 2 with a constraint:
  - Minimize $f(x) = x_1^2 + 4x_2^2 + 8$
    subject to: $g(x) = x_1 + 2x_2 - 4 = 0.$
    - In this case we can solve for x_1 in $g(x)$ to get $x_1 = 4 - 2x_2$.
    - Then f becomes: $8x_2^2 - 16x_2 + 24 = 8\left[(x_2 - 1)^2 + 2\right].$
    - So $x_2 = 1$ and $x_1 = 2$ giving a value for f that is 16.



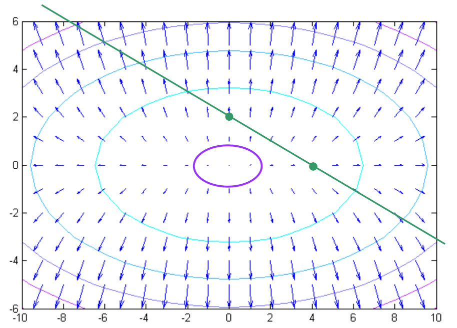
$g(x) = x_1 + 2x_2 - 4 = 0$

Problems with Constraints (2)

- In the last problem we were "lucky" because $g(x_1,x_2) = 0$ could be solved for x_1 or x_2 so that a substitution could be made into $f(x_1,x_2)$ reducing it to a minimization problem in one variable.
- In general, we cannot rely on this strategy.
  - The constraint $g(x_1,x_2) = 0$ could be so complicated that it is impossible to solve for one of its variables.
- Note that the previous visualization in "f vs x space" cannot be carried over to Problem 3.
- We need a new approach.
  - This will require level curves, gradients, and eventually the Lagrange multiplier strategy.

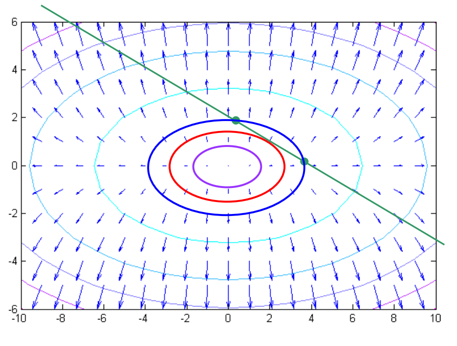
A Different Approach (1)

- Consider viewing the f vs x scenario from above, along the f axis: We would see a series of level sets:
  - The constraint $g(x) = x_1 + 2x_2 - 4 = 0$ appears as the green line.
  - There are 3 cases to analyze when considering the relationship between a level set and the constraint.
  - Case 1: For low values of r the level set does not meet the constraint. So, even though the f values are very small we must reject these points because they do not satisfy the constraint.



A Different Approach (2)

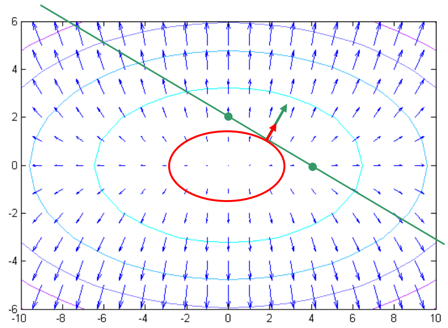
- Case 2: As r increases the level set will eventually just meet the constraint in a tangential fashion. This is the red ellipse.
- Case 3: As r continues to increase the level set will cut the constraint in two places as illustrated by the blue ellipse.



  These two points satisfy the constraint but we can do better! There are points between them on the constraint that correspond to intersections of the constraint with level sets of f having lower r values.
- So, Case 2 is the situation that we want to discover.

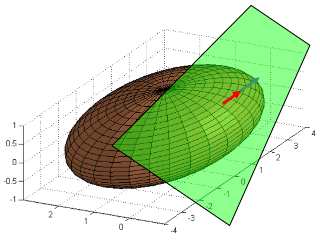
A Different Approach (3)

- Case 2 is characterized by having the level curve of $f(x_1,x_2)$ tangent to the constraint curve $g(x_1,x_2) = 0$.
- In other words: The normal to both f and g at the point of contact are in the same direction.



A Different Approach (4)

- Notice that this visualization will also work for our Problem 3:
- In this case the level sets are ellipsoids that keep expanding until they meet the constraint surface.
- As before we want the normal to the constraint surface to be parallel to the normal of the level surface.



The Lagrange Formulation (1)

- Parallel normals at the point p can be expressed as:

$$\nabla f(p) = \lambda \nabla g(p)$$
$$g(p) = 0.$$

That is: The gradient of f at p is a constant multiple of the gradient of g at p.

  - The Lagrange formulation states:

To find the extreme points of a function $f(x) = f(x_1, x_2, \ldots, x_n)$ that are also subject to the constraint that they lie on $g(x) = g(x_1, x_2, \ldots, x_n) = 0$ we form the Lagrangian

$$L(x, \lambda) = f(x) - \lambda g(x)$$

and find points x such that $\nabla L(x, \lambda) = 0.$

- Such a point p will be one of the required extreme points.

The Lagrange Formulation (2)

- Recall: $L(x, \lambda) = f(x) - \lambda g(x).$
- Note the significance of $\nabla L(x, \lambda) = 0.$
  - To compute the gradient of L we are taking partial derivatives with respect to all the x_i and also λ.
  - This gives us:

$$\left.\begin{array}{l} \dfrac{\partial L}{\partial x_i} = 0 \quad \Rightarrow \quad \dfrac{\partial f}{\partial x_i} = \lambda \dfrac{\partial g}{\partial x_i} \quad \forall i = 1, 2, \ldots, n \\ \dfrac{\partial L}{\partial \lambda} = 0 \quad \Rightarrow \quad g(x) = 0. \end{array}\right\}$$

Now solve $n + 1$ equations in $n + 1$ unknowns.

  - The first set of equations is actually stating:

$$\nabla f(x) = \lambda \nabla g(x).$$

Importance of the Lagrange Formulation

- Recall the requirement:

  $$\nabla f(x) = \lambda \nabla g(x) \qquad g(x) = 0$$

  when x is an extreme point p.

  - The previous slide shows that this is equivalent to

    $\nabla L(x,\lambda) = 0$ with $L(x,\lambda) = f(x) - \lambda g(x)$.

  - Why is this important?
    - Finding critical points of $\nabla L(x,\lambda) = 0$ is a problem that does not mention the constraint as a separate issue.
    - In that sense our problem is simpler.
    - The tradeoff is that we are now dealing with a higher dimensional space ($n + 1$ instead of n dimensions).

Constrained Problem 2 à la Lagrange

- Recall Problem 2 with a constraint:
  - Minimize $f(x) = x_1^2 + 4x_2^2 + 8$
    subject to: $g(x) = x_1 + 2x_2 - 4 = 0.$
  - The Lagrangian is:

    $$L(x,\lambda) = f(x) - \lambda g(x) = x_1^2 + 4x_2^2 + 8 - \lambda(x_1 + 2x_2 - 4).$$

  - Setting $\nabla L(x,\lambda) = 0$ gives:

$$\left.\begin{array}{l} \left.\begin{array}{l} \dfrac{\partial L}{\partial x_1} = 2x_1 - \lambda = 0 \\ \dfrac{\partial L}{\partial x_2} = 8x_2 - 2\lambda = 0 \end{array}\right\} \Rightarrow x_1 = 2x_2 \\ \dfrac{\partial L}{\partial \lambda} = -(x_1 + 2x_2 - 4) = 0. \end{array}\right\} \Rightarrow x_1 = 2 \quad x_2 = 1$$

the same solution we saw earlier.

Extra Constraints (1)

- For some applications we will have several constraints.
  - To motivate this, we go back over our three problems this time imposing *2* constraints, call them $g_1(x) = 0$ and $g_2(x) = 0$.
    - Problem 1: Both constraints $g_1(x) = 0$ and $g_2(x) = 0$ would have to have at least one point on the real axis that simultaneously satisfied these conditions (very unlikely!).
      - We would then consider the value of f at this point or these points, if they exist.
      - This variation of problem 1 is essentially over-constrained and the analysis will usually come up with the result that there is no solution.

Extra Constraints (2)

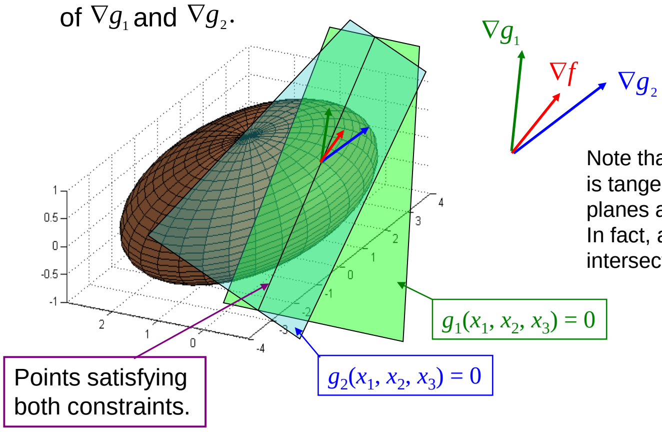
- Problem 2: The enforcement of both constraints $g_1(x_1, x_2) = 0$ and $g_2(x_1, x_2) = 0$ leads to a set of discrete points in the (x_1, x_2) plane where the constraints are satisfied.
  - We would then consider the value of f at this point or these points.
  - While not over-constrained, this variation of problem 2 is not very interesting because of the discrete nature of the investigation (as in Problem 1 with one constraint).
- Problem 3: The enforcement of the two constraints $g_1(x_1, x_2, x_3) = 0$ and $g_2(x_1, x_2, x_3) = 0$ leads to a curve containing points in the (x_1, x_2, x_3) plane where the constraints are satisfied.
  - Now that our feasible set contains an infinite number of points we are back to an interesting situation.
  - The curve is a 2D continuum that is the intersection of $g_1 = 0$ and $g_2 = 0$.

Extra Constraints (3)

- Problem 3 Recap: The enforcement of the two constraints $g_1(x_1, x_2, x_3) = 0$ and $g_2(x_1, x_2, x_3) = 0$ leads to a curve containing points in the (x_1, x_2, x_3) plane where the constraints are satisfied.
  - Intuitively, we would want to increase the level set parameter r until the level surface contacts the intersection curve in a tangential fashion.
  - But how do we express this requirement in terms of the gradient of f?
  - Asking for points p such that ∇f is parallel to *both* ∇g_1 and ∇g_2 is a "show stopper" because ∇g_1 and ∇g_2 could be different from each other for all p (consider two intersecting planes).
  - Going back to the single constraint case, here is another way to express the requirement that gradient of f is parallel to the gradient of g: We ask that ∇f is in the subspace defined by ∇g.
  - This immediately generalizes for our immediate needs: We ask that ∇f lies in the subspace defined by linear combinations of ∇g_1 and ∇g_2.

Extra Constraints (4)

- Problem 3 Recap: The enforcement of the two constraints $g_1(x_1, x_2, x_3) = 0$ and $g_2(x_1, x_2, x_3) = 0$ leads to a curve containing points in the (x_1, x_2, x_3) plane where the constraints are satisfied.
  - We ask that ∇f lies in the subspace defined by linear combinations of ∇g_1 and ∇g_2.



Note that the line of intersection is tangent to the ellipsoid but the planes are not necessarily tangent. In fact, at least one plane intersects the ellipsoid.

Note that three constraints: g_1=0, g_2=0, g_3=0 for Problem 3 leads to the discrete case again.

Extra Constraints: Lagrange Formulation

- Our previous requirement is formalized as:

$$\nabla f(x) = \sum_{i=1}^{k} \lambda_i \nabla g_i(x)$$
$$g_i(x) = 0 \quad \forall i = 1, 2, \dots k.$$

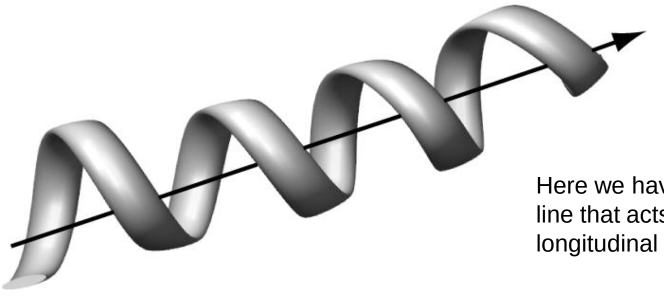
  - As before x is really $(x_1, x_2, \dots, x_n)^T$.
  - To avoid the over-constrained or discrete points case we assume that the number of constraints $k < n$.
  - This gives the Lagrangian: $\boxed{L(x,\lambda) = f(x) - \sum_{i=1}^{k} \lambda_i g_i(x)}$
  - Setting $\nabla L(x,\lambda) = 0$ leads to $n + k$ equations in $n + k$ unknowns.

Extra Constraints: Lagrange Formulation

- Summary:
  - To find extreme points of $f(x)$ subject to the constraints $g_i(x) = 0$ for $i=1,2,\dots,k$ with $k < n$:
    - (the notation x represents $(x_1, x_2, \dots, x_n)^T$)
  - We form the Lagrangian: $L(x,\lambda) = f(x) - \sum_{i=1}^{k} \lambda_i g_i(x)$ where $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)^T$.
  - Then the extreme points are calculated by solving the $n + k$ equations in $n + k$ unknowns derived from $\nabla L(x,\lambda) = 0$.

Inertial Axes (1)

- An application of Lagrange multipliers:
  - Very often in a structural analysis, we want to approximate a secondary structural element with a single straight line.



Here we have a straight line that acts as the longitudinal axis of a helix.

Inertial Axes (2)

- The diagram gives us an indication of a reasonable strategy for the determination of the axis: We should have the straight line positioned among the atoms so that it is closest to all these atoms in a least squares sense.
  - Our objective is to find this “best” helix axis.
  - A reasonable strategy would be to have the axis pass through the centroid of all the atoms with a direction chosen to minimize the sum of the squares of the perpendicular distances from the atoms to the helix axis.

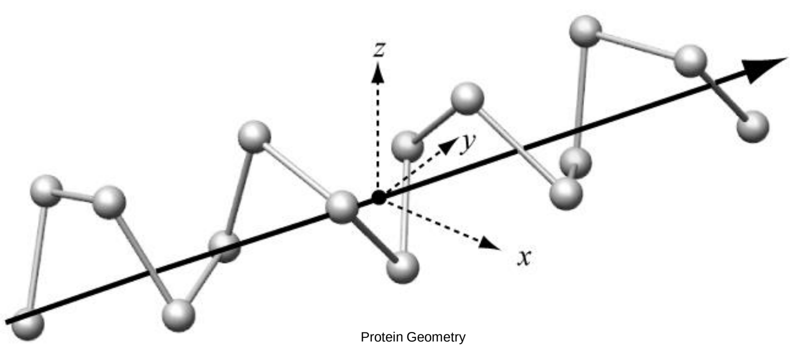
Inertial Axes (3)

- Mathematical derivation of the axes:
  - If $\|d_i\|$ represents the perpendicular distance between atom $a^{(i)}$ and the helix axis, then we chose the axis direction so that it minimizes the sum:

$$S = \sum_{i=1}^{N} \|d_i\|^2$$

Inertial Axes (4)

- It should be stated that such an axis could also be computed to go among the atoms of a beta strand, in fact, any arbitrary set of atoms.

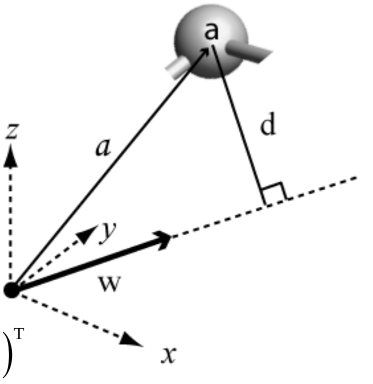


Inertial Axes (5)

- We now provide an analysis that shows how the axis is dependent on the coordinates of the chosen atoms.
  - To make the analysis simpler, we first consider one atom, call it a, with position vector $a = (a_x, a_y, a_z)^{\mathrm{T}}$.
  - As noted earlier, this vector is using a frame of reference that is an x, y, z coordinate system with origin positioned at the centroid of the atoms under consideration (see previous figure).

Inertial Axes (6)

- Analysis continued:
  - The perpendicular distance from the atom to the axis is illustrated in the figure.
  - We are trying to determine the values of w_x, w_y, and w_z that define the axis direction.
  - These are the components of a unit vector w and so $w = (w_x, w_y, w_z)^{\mathrm{T}}$ with $w^{\mathrm{T}} w = 1$.



The required axis is a scalar multiple of w and d represents the perpendicular vector going from atom a to this axis.

Inertial Axes (7)

- Analysis continued:
  - Some trigonometry defines the square of the norm of d as:

$$\|d\|^2 = \|a\|^2 \sin^2\theta = \|a\|^2\left[1-\cos^2\theta\right] = \|a\|^2\left[1-\frac{\left(a^{\mathrm{T}}w\right)^2}{\|a\|^2\|w\|^2}\right]$$

This can be written as: $\|w\|^2$

This = 1.

  - So:

$$\|d\|^2 = \|a\|^2\cdot 1-\left(a^{\mathrm{T}}w\right)^2 = \|a\|^2\|w\|^2-\left(a^{\mathrm{T}}w\right)^2$$

This expression for d is quadratic in w.

Inertial Axes (8)

- Analysis continued:
  - Now we replace a and w with their coordinate representations:

$$\begin{aligned}\|d\|^2 &= \left(a_x^2+a_y^2+a_z^2\right)\left(w_x^2+w_y^2+w_z^2\right)-\left(a_xw_x+a_yw_y+a_zw_z\right)^2\\ &= Aw_x^2+Bw_y^2+Cw_z^2-2\left[Fw_yw_z+Gw_zw_x+Hw_xw_y\right]\end{aligned}$$

where:

$$\begin{aligned} A &= a_y^2+a_z^2 & F &= a_ya_z\\ B &= a_z^2+a_x^2 & G &= a_za_x\\ C &= a_x^2+a_y^2 & H &= a_xa_y. \end{aligned}$$

Inertial Axes (9)

- Analysis continued:
  - Finally, an elegant representation for d:

$$\|d\|^2 = \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}^{\mathrm{T}} \begin{bmatrix} A & -H & -G \\ -H & B & -F \\ -G & -F & C \end{bmatrix} \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}.$$

  - Now recall that we want to minimize S which is the sum of **all** such squared norms going across all N atoms in the set.

Inertial Axes (10)

- Analysis continued:
  - We let the coordinates of the i^{th} atom a$^{(i)}$ be represented by $a^{(i)} = \left(a_x^{(i)}, a_y^{(i)}, a_z^{(i)}\right)^{\mathrm{T}}$, then

$$S = \sum_{i=1}^{N} \|d_i\|^2 = \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}^{\mathrm{T}} \begin{bmatrix} D_{xx} & -E_{xy} & -E_{xz} \\ -E_{yx} & D_{yy} & -E_{yz} \\ -E_{zx} & -E_{zy} & D_{zz} \end{bmatrix} \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}$$

where:

$$D_{xx} = \sum_{i=1}^{N} \left(a_y^{(i)}\right)^2 + \left(a_z^{(i)}\right)^2 \qquad E_{yz} = E_{zy} = \sum_{i=1}^{N} a_y^{(i)} a_z^{(i)}$$

$$D_{yy} = \sum_{i=1}^{N} \left(a_z^{(i)}\right)^2 + \left(a_x^{(i)}\right)^2 \qquad E_{xz} = E_{zx} = \sum_{i=1}^{N} a_z^{(i)} a_x^{(i)}$$

$$D_{zz} = \sum_{i=1}^{N} \left(a_x^{(i)}\right)^2 + \left(a_y^{(i)}\right)^2 \qquad E_{xy} = E_{yx} = \sum_{i=1}^{N} a_x^{(i)} a_y^{(i)}.$$

Inertial Axes (11)

- Analysis continued:
  - Finally, we have the quantity that we want to minimize expressed as a function of w and the coordinates of the atoms:

$$\boxed{S = w^{\mathrm{T}} T w}$$

where:

$$T = \begin{bmatrix} D_{xx} & -E_{xy} & -E_{xz} \\ -E_{yx} & D_{yy} & -E_{yz} \\ -E_{zx} & -E_{zy} & D_{zz} \end{bmatrix}$$

Physicists call this the **inertial tensor**.

Inertial Axes (12)

- Analysis continued:
  - We can now formulate our minimization problem as a Lagrange multiplier problem in which we minimize S subject to the constraint that $w^{\mathrm{T}} w = 1$.
    - The Lagrangian will be:

$$L(w,\lambda) = w^{\mathrm{T}} T w - \lambda\left(w^{\mathrm{T}} w - 1\right)$$

where λ is the Lagrange multiplier.

- Setting the usual derivatives to zero:

$$\frac{\partial L}{\partial w} = 2Tw - 2\lambda w = 0$$

Inertial Axes (13)

- Analysis finished:
  - We now recognize this as an eigenvector problem:

$$Tw = \lambda w$$

  - The 3 by 3 tensor matrix T is symmetric and since S is always positive for any vector w we see that T is also positive definite.
    - Consequently, all three eigenvalues are positive.
  - Note that once we have an eigenvalue, eigenvector pair we can write:

$$S = w^{\mathrm{T}}Tw = \lambda w^{\mathrm{T}}w = \lambda.$$

Inertial Axes (14)

- So, we have the solution to our problem.
  - Once T is calculated, we compute its eigenvalues and the required w (specifying the directions for our axis) will be the eigenvector corresponding to the **smallest** eigenvalue.
    - Recall that we wanted to **minimize** S.

Inertial Axes (15)

- Significance of three eigenvalues:
  - It should also be noted that there are three eigenvectors produced by this procedure and because of the symmetry of T, they are mutually orthogonal and can be used as an orthogonal basis for the set of atoms under consideration.
    - In some applications, this *inertial frame of reference* is a useful construct and the eigenvectors are called the principle axes of inertia.