ECH4905 - Special Topics in ChemE - Optimization

Andres Espinosa

January 23, 2025

Contents

1 Chapter 1 - Optimization Models

1.1 Tuesday 01/14/2025

1.1.1 Modeling Overview

An optimization problem is represented by degrees of freedom that inform decisions. The degrees of freedom of a problem dictate the number of decisions that can be made.

- n Decisions $> 0 \implies$ optimization problem
- n decisions $= 0 \implies$ simulation

The number of decisions is equal to the degrees of freedom, which is equivalent to the number of variables minus the number of inequalities.

1.1.2 Example Problem

This example problem will use a toy example of minimizing the perimeter of a rectangle with side lengths x and y.

$$\begin{aligned} \text{minimize} \quad & P && (1) \\ \text{subject to} \quad & xy = 2000 && (2) \\ & P = 2x + 2y && (3) \\ & x \geq 0, y \geq 0 && (4) \end{aligned}$$

with variables x, y, P. Since there are two equalities in the optimization problem above and three variables, there is one degree of freedom. This can be seen more plainly by reformulating the problem into the below problem

$$\begin{aligned} \text{minimize} \quad & 2(\frac{2000}{y}) + 2y && (5) \\ \text{subject to} \quad & y \geq 0 && (6) \end{aligned}$$

1.1.3 General Optimization Definitions

A general optimization problem can be written as

$$\begin{aligned} \text{minimize} \quad & F(x, y) && (7) \\ \text{subject to} \quad & h(x, y) = 0 && (8) \\ & g(x, y) \leq 0 && (9) \end{aligned}$$

where $x \in \mathbb{R}^n, y \in \{0, 1\}^m$. This general optimization problem can be broken down into different problems with levels of complexity.

Firstly, a **linear program** (LP) is defined as

$$\begin{aligned} \text{minimize} \quad & F(x) && (10) \\ \text{subject to} \quad & h(x) = 0 && (11) \\ & g(x) \leq 0 && (12) \end{aligned}$$

Where $x \in \mathbb{R}^n$ and F, h, g are affine functions. Linear problems typically make many approximations about the real physics/constraints that are in the world. These approximations are a tradeoff with computational tractability. The Simplex method is typically used to solve linear programs.

Secondly, a **non-linear program** (NLP) is defined as

$$\begin{aligned} \text{minimize} \quad & F(x) && (13) \\ \text{subject to} \quad & h(x) = 0 && (14) \\ & g(x) \leq 0 && (15) \end{aligned}$$

Where $x \in \mathbb{R}^n$ and any of the functions F, h, g are non-linear. A subset of NLPs are those problems that are *convex*. Convexity in a non-linear program is achieved when F, g are convex functions and h is an affine function. Some algorithms that are used to solve NLPs are SQP, IPOPT.

A higher degree of complexity, are **mixed integer linear programs** (MILP) defined as

$$\begin{aligned} \text{minimize} \quad & F(x,y) && (16) \\ \text{subject to} \quad & h(x,y) = 0 && (17) \\ & g(x,y) \leq 0 && (18) \end{aligned}$$

where $x \in \mathbb{R}^n, y \in \{0,1\}^m$. and F, h, g are affine functions which can be represented as $F(x,y) = Ax + By$. These problems have applications in chemical engineering with examples such as biomass, waste, and fuel supply chains. In a biomass biorefinery location problem, you would have a set of locations to pick from y and you want to minimize the cost. An algorithm used to solve these problems is branch-and-bound.

With the highest degree of complexity, a **mixed integer non-linear program** (MINLP) is defined as

$$\begin{aligned} \text{minimize} \quad & F(x,y) && (19) \\ \text{subject to} \quad & h(x,y) = 0 && (20) \\ & g(x,y) \leq 0 && (21) \end{aligned}$$

where $x \in \mathbb{R}^n, y \in \{0,1\}^m$ and any of the F, h, g functions are non-linear. BARON, SCIP, MAINGO are algorithms that have been developed in the last 10 years in response to solve MINLPs. Chemical engineering is naturally non-linear and there has been a big push from the chemical optimization community to create solutions for MINLPs. These are complex problems but also important problems.

1.1.4 Specific Model Criteria

Optimization problems can be also defined by the **number of objectives** they have.

- Single objective
- Multiple objective

Multiple objectives introduce a tradeoff between different objectives with each other. If the number of objectives is low, then it is easy to identify and visualize the tradeoffs. The concept of pareto optimal will be explored further in the course.

$$\text{minimize} \quad \langle F_1(x), F_2(x), \ldots, F_k(x) \rangle \tag{22}$$

Optimization problems can also have a level of **uncertainty** in the problem. For example, the price and cost parameters may be random and uncertain. The majority of time, the randomness and uncertainty in models are ignored by modelers. A good guess is used and substituted for the random parameters such as an average or other estimate. However, it is possible to create stochastic models that attempt to handle uncertainty.

- Deterministic models: Uncertainty is ignored and stochastic parameters are replaced by a fixed "good guess".
- Stochastic models: Uncertainty is handled and incorporated in the constraints.

Typically, stochastic models are significantly more difficult to compute than deterministic models. So, a deterministic model should be used and solved before a stochastic one is approached.

Optimization models are also specified by their **type/number of constraints**

- Unconstrained
- Constrained

1.1.5 Optimization Concepts

For general optimization problem defined above, a **feasible region** is defined as the set of $\langle x, y \rangle$ that satisfy all equalities and inequalities.

$$\begin{aligned}
\text{minimize} \quad & -xy && (23)\\
\text{subject to} \quad & x \geq 0, x \leq 1 && (24)\\
& y \geq 0, y \leq 1 && (25)
\end{aligned}$$

For this problem, the feasible region R is

$$R(x, y) = \{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq 1\} \tag{26}$$

A **local minimum** with respect to a distance ε is defined as a point (x^*, y^*) that satisfies $F(x^*, y^*) \leq F(x, y), \|x^* - x\| \leq \varepsilon$. A **global minimum** is a point that satisfies $F(x^*, y^*) \leq F(x, y)$.

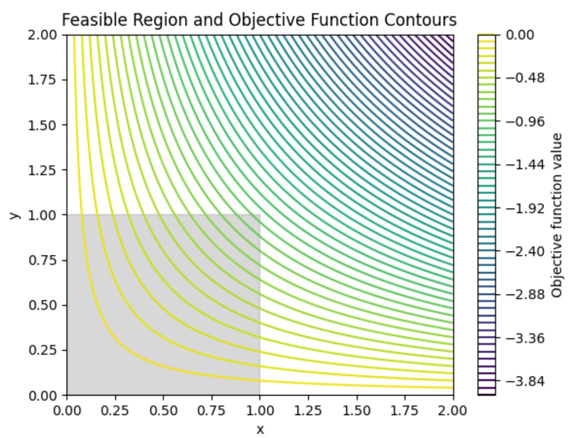

Figure 1: A plot of the optimization problem $F(x, y) = -xy$

2 Chapter 2 - Mathematics Review

2.1 Thursday 01/16/2025

2.1.1 Chapter Math Outline

The chapters over the next few lessons will include some math that will be used in different chapters.

- Chapter 1: Linear algebra

  Matrices

  Eigenvalues

- Chapter 2: Convex Analysis

  Convexity

  Quadratic forms

  Taylor Series

2.1.2 Linear Algebra Review

A vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{27}$$

This vector represents a direction and magnitude in n dimensions.
The $l2$-norm of a vector is defned as

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \dots x_n^2} \tag{28}$$

Vector addition is defined as

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \tag{29}$$

Scalar multiplication of a scalar $a \in \mathbb{R}$ with a vector $\mathbf{x} \in \mathbb{R}^n$

$$a\mathbf{x} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} \tag{30}$$

The dot product of two vectors $\mathbf{x}, \mathbf{y}, \in \mathbb{R}^n$ is defined as

$$\mathbf{x}^\top \mathbf{y} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{31}$$

With equivalent notation $\mathbf{x}^\top\mathbf{y} = \langle\mathbf{x},\mathbf{y}\rangle$

$$\|\mathbf{x}-\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\|\mathbf{x}\|_2\|\mathbf{y}\|_2\cos\theta \tag{32}$$
$$\|\mathbf{x}-\mathbf{y}\|_2^2 = (\mathbf{x}-\mathbf{y})^\top(\mathbf{x}-\mathbf{y}) \tag{33}$$

The Cauchy-Schwarts Inequality is derived by the following

$$\cos\theta = \frac{x^\top y}{\|x\|\|y\|} \tag{34}$$
$$\frac{x^\top y}{\|x\|\|y\|} \leq 1 \tag{35}$$
$$x^\top y \leq \|x\|\|y\| \tag{36}$$

The triangle inequality is defined as

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{37}$$

A matrix $A \in \mathbb{R}^{m\times n}$ is defined as

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \tag{38}$$

The transpose of a matrix $A^\top$ flips each value for the row and column as such

$$A^\top = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} \tag{39}$$

Matrix addition is element-wise and can be shown as such between two matrices $A, B \in \mathbb{R}^{m\times n}$

$$A + B = \begin{bmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & \dots & \vdots \\ a_{m1}+b_{m1} & \dots & a_{mn}+b_{mn} \end{bmatrix} \tag{40}$$

Scalar multiplication of a matrix A with a scalar α is defined as

$$\alpha A = \begin{bmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & \dots & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{bmatrix} \tag{41}$$

Matrix multiplication between two matrices $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$ is defined as

$$AB = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \dots & b_{1p} \\ \vdots & \dots & \vdots \\ b_{n1} & \dots & b_{np} \end{bmatrix} \tag{42}$$

With the following properties

- $AB \neq BA$
- $(ABC)^\top = C^\top B^\top A^\top$

The inverse of a matrix has the following properties

- $AA^{-1} = I$
- $(AB)^{-1} - B^{-1}A^{-1}$
- $(B^\top)^{-1} = (B^{-1})^\top$
- $(A^{-1})^{-1} = A$

Orthonormal matrices have the properties

- $O^\top O = I$
- $O^\top = O^{-1}$

Some following matrix partitions are useful

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} + \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_1 + A_1 \\ B_1 + B_2 \end{bmatrix} \tag{43}$$

$$\begin{bmatrix} A \\ B \end{bmatrix}^\top = \begin{bmatrix} A^\top & B^\top \end{bmatrix} \tag{44}$$

$$A \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}^\top = \begin{bmatrix} AB_1 & AB_2 \end{bmatrix} \tag{45}$$

The determinant of a matrix $A \in \mathbb{R}^{n \times n}$ is calculated by subtracting the product of the diagonals of a matrix recursively. The minor of a matrix is the section of a matrix that is achieved when removing a section. Simply, the determinant of a matrix can also be defined as the product of the eigenvalues.

$$\det A = \prod \lambda \tag{46}$$

2.1.3 Spaces

A vector space is defined as a set of all vectors with some properties. We define a vector space $\mathbb{V}$

$$u + v \in \mathbb{V} \tag{47}$$
$$u + (v + w) = (u + v) + w \tag{48}$$
$$u + v = v + u \tag{49}$$

A linearly dependent system of vectors $v_i \in \mathbb{R}^n$ satifies the following for a set of scalars c_i

$$c_1 v_1 + \cdots + c_m v_m = 0 \tag{50}$$

The span of a set of vectors v_i is the set of all vectors that can be created with a linear combination of those vectors.

$$\{x | c_1 v_1 + \cdots + c_m v_m = x\} \tag{51}$$

The basis of a space is the minimum number of vectors needed to span a vector space.

2.2 Tuesday 01/21/2025

2.2.1 Matrix eigenvalues

We define λ as the eigenvalues of a matrix. We investigate what it means for matrices with different eigenvalues.

- $\lambda_i \geq 0 \to$ Matrix A is positive semi-definite
- $\lambda_i > 0 \to$ Matrix A is positive definite
- $\lambda_i \leq 0 \to$ Matrix A is negative semi-definite
- $\lambda_i < 0 \to$ Matrix A is negative definite
- $\lambda_i < 0, \lambda_i > 0 \to$ Matrix A is indefinite

2.2.2 Convex Sets

A convex set is defined as a set where the line through all points is contained in the same set.

Half Spaces:
Half spaces are the space under a line, defined with the parameters $c \in \mathbb{R}^n z \in \mathbb{R}$,

$$\{x | c^\top x \leq z, x \in \mathbb{R}^n\} \tag{52}$$

A half space is open if the inequality is strict.
The intersection of a finite number of closed half spaces is known as a polytope. If a polytope is bounded on all sides, it is also known as a polyhedron.

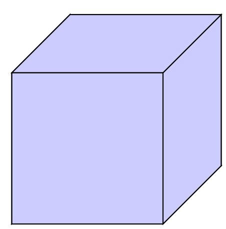

Figure 2: A polyhedron

Extreme Points:
Consider a convex set $C \subseteq \mathbb{R}^n$. Let $z \in \mathbb{R}^n$, z is an extreme point of C if $z \in C$ and there are no $x, y \in C$ such that $z = (1 - \lambda)x + \lambda y$. In other words, the point z is not on a line defined by two points. The figure 3 shows a polyhedron. It is impossible to represent the red points as a positive combination of two other points in this square. Therefore, they are extreme points.

Convex combinations:
Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ be a set of m vectors in $\mathbb{R}^n$. A convex combination of these vectors is a point of the form

$$\lambda_1 \mathbf{x}_1 + \cdots + \lambda_m \mathbf{x}_m \tag{53}$$

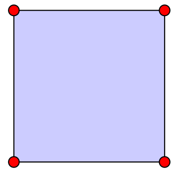

Figure 3: A 2D polyhedron with emphasized vertices

where $\mathbf{1}^\top \lambda = 1, \lambda \succeq 0$

A Simplex:
A simplex is the simplest possible polytope in a given dimension. Figure 4 shows the simplexes in 0 to 3 dimensions.

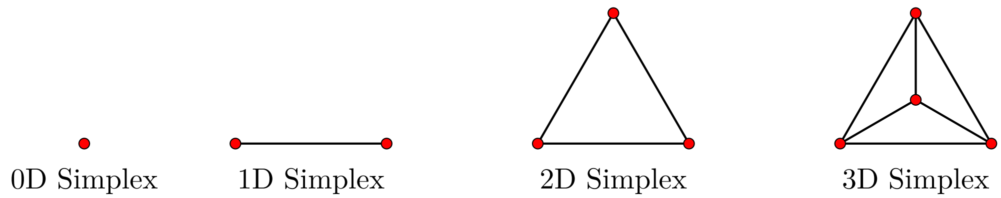

Figure 4: Simplexes in 0D to 3D

Convex Hull
The convex hull of a set S is defined as the smallest convex set that contains S. It is also defined as the intersection of all convex sets which contain S. Figure 5 shows the convex hull of a nonconvex set.

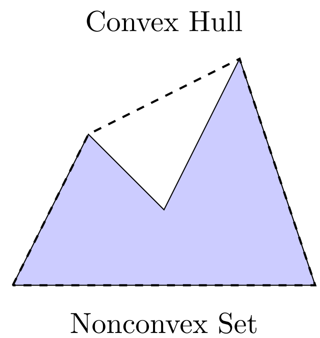

Figure 5: A 2D nonconvex set and its convex hull

Farkas Theorem:
The Farkas Theorem is a theorem that allows for a certificate of feasibility (or infeasibility) of an optimization problem. With the variables $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^n$, exactly one of the following two propositions is true.

- $Ax = b, x \succeq 0$ for some $x \in \mathbb{R}^n$. In other words, b is in the cone spanned by convex combinations of the columns of A.

- $A^\top y \succeq 0, b^\top y \leq 0$, for some $y \in \mathbb{R}^m$. In other words, the angle between y and b is greater than 90 and the angle between y and every column of A is less than 90.

$$Ax = b \rightarrow \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{54}$$

2.2.3 Convex Functions

Derivatives:
Gradient refresher

$$\nabla f(x) = \begin{bmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \\ \vdots \\ \frac{df}{dx_n} \end{bmatrix} \tag{55}$$

A directional derivative is the transpose of the gradient times a direction $\nabla f(x)^\top d$. Consider a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that is differentiable. A direction is a descent direction if $d^\top \nabla f(x) < 0$. The jacobian of a function is a generalized derivative for a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. It is defined as

$$J(x) = \begin{bmatrix} \nabla f_1(x)^\top \\ \nabla f_2(x)^\top \\ \vdots \\ \nabla f_m(x)^\top \end{bmatrix} \tag{56}$$

The hessian of a function is the second derivative of a function. The Hessian is symmetric as long as the mixed derivatives are equal to each other.

$$H(f(x)) = \begin{bmatrix} \frac{d^2 f}{dx_1^2} & \cdots & \frac{d^2 f}{dx_1 dx_n} \\ \vdots & \vdots & \vdots \\ \frac{d^2 f}{dx_n dx_1} & \cdots & \frac{d^2 f}{dx_n^2} \end{bmatrix} \tag{57}$$

Convex Functions:
Let C be a convex subset of $\mathbb{R}^n$, and f(x) be a real-valued function defined on C. The function f is convex if Jensens inequality below holds. A function is strictly convex if the below property holds with strict inequality.

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \tag{58}$$

In words, a function is convex if the line between any two points lies above every point of the function between those two points. Figure 6 shows a convex function and how the line between any two points lies above the graph. Another way to define convex functions is with respect to the Hessian of the function $H(f(x))$. A function is convex if the Hessian of the function is positive semi definite $H(f(x)) \succeq 0$. Convexity can also be defined in sections of a function. For example, let S

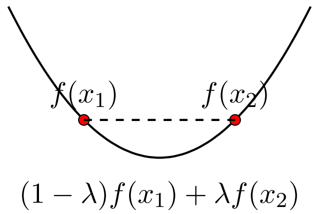

Figure 6: A convex function and the line between two points

be anon-empty open set in $\mathbb{R}^n$. If $f(x_0)$ is convex at the point x_0, then H is positive semi-definite at that point.

Example Problem:
Determine if the following equation is convex

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 \tag{59}$$

$$\nabla f(x) = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_2 + 2x_1 \\ 2x_3 \end{bmatrix} \tag{60}$$

$$\nabla^2 f(x) = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{61}$$

Calculating the eigenvalues

$$\det \begin{bmatrix} 2-\lambda & 2 & 0 \\ 2 & 2-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{bmatrix} \tag{62}$$

$$(2-\lambda)(2-\lambda)(2-\lambda) - 2(2)(2-\lambda) + 0 = 0 \tag{63}$$
$$(2-\lambda)^3 - 4(2-\lambda) = 0 \tag{64}$$
$$(2-\lambda)((2-\lambda)^2 - 4) = 0 \tag{65}$$
$$\lambda = [2, 0, 4] \tag{66}$$

$\lambda = [0, 2, 4] \succeq 0$, therefore the hessian is positive semi-definite and the function is convex.

2.2.4 Properties of Convex Functions

- if f_i are convex, $\sum_i f_i(x)$ is convex
- if f is convex, $\lambda f(x)$ is convex, where $\lambda isascalar$
- Let f be convex, and g be an increasing function. The convex function $g(f(x))$ is also convex.

2.2.5 Quadratic Forms

The quadratic form of a vector $x \in \mathbb{R}^n$ with parameters $Bin\mathbb{R}^{n\times n}, a \in \mathbb{R}^n, c \in \mathbb{R}$ is

$$f(x) = \frac{1}{2}x^\top Bx + a^\top x + c \tag{67}$$

2.3 Thursday 01/23/2025

2.3.1 Eigenvalues and Eigenvectors

A non-singular matrix is a matrix A that follows the property $\det A \neq 0$.

Eigenvalues:
In order to calculate the eigenvalues of a matrix A, we solve the equation $\det(A - \lambda I) = 0$. For example, we take the matrix

$$A = \begin{bmatrix} 13 & -4 \\ -4 & 7 \end{bmatrix} \tag{68}$$

We subtract λ from the diagonals to get

$$A = \begin{bmatrix} 13-\lambda & -4 \\ -4 & 7-\lambda \end{bmatrix} \tag{69}$$

$$= (13-\lambda)(7-\lambda) - 16 = 0 \tag{70}$$

$$\lambda_1 = 15, \lambda_2 = 5 \tag{71}$$

Not all eigenvalues are always real. The number of eigenvalues is the same as the dimension of the matrix but is not always real.

Eigenvectors
To find the eigenvectors ν of the matrix, we solve $(A - \lambda I)\nu = 0$.

$$A - \lambda I = (\begin{bmatrix} 13 & -4 \\ -4 & 7 \end{bmatrix} - 15 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{72}$$

$$= \begin{bmatrix} -2 & -4 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{73}$$

$$-2v_1 - 4v_2 = 0 \tag{74}$$

$$v_1 + 2v_2 = 0 \tag{75}$$

$$v_1 = -2v_2 \tag{76}$$

$$\text{Infinite solutions, as long as the vector looks like } \begin{bmatrix} -2a \\ a \end{bmatrix} \tag{77}$$

Now we use the other eigenvalue $\lambda = 5$

$$A - \lambda I = (\begin{bmatrix} 13 & -4 \\ -4 & 7 \end{bmatrix} - 5 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}) \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{78}$$

$$= \begin{bmatrix} 8 & -4 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{79}$$

$$\text{Infinite solutions, as long as the vector looks like } \begin{bmatrix} a/2 \\ a \end{bmatrix} \tag{80}$$

2.3.2 Convexity in Functions

A bilinear function is a function in the form

$$f(x_1, x_2) = x_1 x_2 \tag{81}$$

The hessian of this function looks like

$$H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{82}$$

The eigenvalues of this matrix are $(-\lambda)^2 - 1 = 0, \lambda = \pm 1$. This matrix is indefinite and therefore non-convex. This function creates a saddle and is non-convex in both variables but is convex in either variable at a fixed falue of the other.