Iordanis Koutsopoulos

NETWORK OPTIMIZATION
LECTURE NOTES

Part 1 [1]

[1] Based on Notes from Fall 2006, and Fall 2010. Organized by Stella Gaki.

Contents

1 Non-linear Programming Problems with Equality Constraints **5**

1.1 Lagrange Theorem for $m = 1$ constraint 7

1.1.1 Definition (as reminder) 8

1.1.2 Example 10

1.2 Lagrange Theorem for $m > 1$ constraints . . . 11

1.2.1 Example 12

1.3 Lagrangian function 13

2 Necessary and Sufficient Conditions **14**

2.1 Sufficient conditions minimum for *convex* functions 15

2.2 Local and global maximum of NLP problems . . 16

2.3 Examples 17

2.3.1 Example 1 17

2.3.2 Example 2 18

2.3.3 Example 3 22

2.3.4 Example 4 23

3 Interpretation of Lagrange multipliers **24**

3.1 Sensitivity analysis for $m > 1$ constraints . . . 26

4 Beamforming **27**

4.1 Starting from an example 27

4.2 Beamforming Basics 28

5 NLP problems with inequality constraints 40

6 Necerrary and Sufficient Conditions 41

6.1 First-order Necessary Conditions for existence of local minimum 41

6.1.1 Karush-Kuhn-Tucker (KKT) theorem . . . 41

6.1.2 Graphical interpretation of KKT theorem . 42

6.1.3 Example 44

6.2 Other forms of optimization problems 46

6.3 Second-order necessary conditions for existence of local minimum 50

6.4 Second Order Sufficient Conditions for existence of local minimum 51

6.5 Second-order Sufficient Conditions for Convex Functions 53

7 Sensitivity analysis 54

7.1 Problem 55

8 The water-filling algorithm 57

9 Lagrangian Duality in NLP 64

9.1 The Lagrangian 64

9.2 The Lagrangian dual function 65

9.3 Lower bounds on optimal value 66

9.4 Lagrangian dual problem 67

9.5 Week Duality 68

9.6 Strong Duality 70
9.7 Solving primal problem (P) using Lagrangian dual (LD) 71

10 Saddle-point interpretation **72**
10.1 Max-min characterization of weak and strong duality 72
10.2 Game theory interpretation of saddle point . . . 74

11 Decomposition theory **75**
11.1 Example : Network pricing 75
11.2 Primal-Dual Algorithm 80

12 Extra Notes **82**

1 Non-linear Programming Problems with Equality Constraints

In this part of the course, we will discuss methods for solving a class of nonlinear constrained optimization problems that can be formulated as:

$$\begin{array}{lll} \min & f(x) & \\ \text{s.t} & h_i(x) = 0, & i = 1, ..., m \\ & g_j(x) \leq 0, & j = 1, ..., p \end{array}$$

where
$\mathbf{x} \in \Re^n$, $f : \Re^n \rightarrow \Re$, $g_j : \Re^n \rightarrow \Re$, and $m \leq n$.

In particular, we will first consider the class of Non-linear Programming problems with constraints that can expressed as equalities $\{h_i(\mathbf{x}) = 0, i = 1, \ldots, m\}$. A point $\mathbf{x}_0$ is called feasible point if it satisfies the constraints.

The constraints $h_i(\mathbf{x}) = 0,\ i = 1, ..., m$ define a surface $\mathcal{S} = \{\mathbf{x} : h_i(\mathbf{x}) = 0,\ i = 1, ..., m\}$.

The tangent plane at a point $\mathbf{x}_0$ on the surface $\mathcal{S}$ is the collection of derivatives at point $\mathbf{x}_0$ of all differentiable curves on $\mathcal{S}$ passing through $\mathbf{x}_0$. A tangent plane to a surface can be visualized as generalizing the tangent line to a point on a curve.

Problem:
We would like to find an explicit characterization of the tangent plane at a point $\mathbf{x}_0$ on the surface defined by the constraints of

our problem,

$$\mathcal{S} = \{\mathbf{x} : h_i(\mathbf{x}) = 0,\ i = 1, ..., m\} \tag{1}$$

as a function of the gradients of the constraint functions h_i, $i = 1, \ldots, m$. For a point $\mathbf{x}_0 \in \mathcal{S}$, we introduce the set of points

$$\mathcal{M} = \left\{\mathbf{y} : \nabla h_i^T(\mathbf{x}_0)\mathbf{y} = 0, \forall i = 1, \ldots, m\right\} \tag{2}$$

that is vector $\mathbf{y}$ denotes all points that are orthogonal to the Gradients of the constaints.

Note that surface $\mathcal{S}$ and $\mathcal{M}$ coicide only when we are working at one or two dimensions. Generally, $\mathcal{S}$ is different from $\mathcal{M}$.

Definition:

A point $\mathbf{x}_0 \in \mathcal{S}$ is said to be a *regular point* if vectors $\nabla h_1(\mathbf{x}_0), \ldots, \nabla h_m(\mathbf{x}_0)$ are linearly independent.

Theorem:

At a regular point $\mathbf{x}_0 \in \mathcal{S} = \{\mathbf{x} : h_i(\mathbf{x}) = 0,\ i = 1, ..., m\}$, the tangent plane is $\mathcal{M}$.

Thus, at regular points we can characterize the tangent plane in terms of the gradients of the constraint functions.

Example:

Consider the surface

$\mathcal{S} = \left\{\mathbf{x} \in \Re^3 : h_1(\mathbf{x}) = x_1 = 0,\ h_2(\mathbf{x}) = x_1 - x_2 = 0\right\}$.

The surface is clearly $\mathcal{S} = \{\mathbf{x} = (0, 0, x_3) : x_3 \in \Re\}$, namely the x_3-axis.

At a point $\mathbf{x}_0 \in \mathcal{S}$, it is

$$\triangledown h_1(\mathbf{x_0}) = \left[\begin{array}{ccc} 1 & 0 & 0 \end{array}\right]$$
$$\triangledown h_2(\mathbf{x_0}) = \left[\begin{array}{ccc} 1 & -1 & 0 \end{array}\right]$$

So, $\nabla h_1(\mathbf{x}_0), \nabla h_2(\mathbf{x}_0)$ are linearly independent $\forall \mathbf{x}_0 \in \mathcal{S}$ and so every point $\mathbf{x}_0 \in \mathcal{S}$ is a reqular point. Then, the tangent plane at $\mathcal{S}$ at point $\mathbf{x}_0$ is:

$$\mathcal{M} = \{\mathbf{y} : \nabla h_i^T(\mathbf{x}_0)\mathbf{y} = 0\}$$

$$= \left\{\mathbf{y} : \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} (y_1 \;\; y_2 \;\; y_3) = 0, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} (y_1 \;\; y_2 \;\; y_3) = 0 \right\}$$

$$\begin{aligned} \Longrightarrow \mathcal{M} &= \{(y_1, \;\; y_2, \;\; y_3) : y_1 = 0, \; y_1 = y_2\} \\ &= \{(0, 0, y_3) : y_3 \in \Re\} \end{aligned} \quad (4)$$

Observe that the surfaces $\mathcal{S}$ and $\mathcal{M}$ coincide.Also, note that $\mathcal{M}$ is not a function of x.That is because, in this specific case, Gradient is not a function of x, too. But, in general cases, $\mathcal{M}$ is dependent on variable x.

1.1 Lagrange Theorem for $m = 1$ constraint

Lemma (for one equality constraint): Let $\mathbf{x_0}$ be a regular point of the surface defined by the equality constraints, $\mathcal{S} =$

$\{h(\mathbf{x}) = 0\}$ and $\mathbf{x}_0$ is a local minimizer of $f : \Re^n \to \Re$, subject to the constraint $h(\mathbf{x}) = 0$. Then all points $\mathbf{y}$ that satisfy $\nabla h(\mathbf{x}_0)^T \mathbf{y} = 0$ also satisfy $\nabla f(\mathbf{x_0})^T \mathbf{y} = 0$. So, that means that both vectors $\nabla h, \nabla f$ are orthogonal to vectors $\mathbf{y}$ on the tangent plane at point $\mathbf{x}_0$ of the surface $\mathcal{S}$ and this means that they are *parallel* to each other. Hence we arrive at the theorem of Lagrange for one constraint $m = 1$ which is stated as follows:

Lagrange Theorem for $m = 1$ constraint: Let the point $\mathbf{x}_0$ be a local minimizer of $f : \Re^n \to \Re$ subject to the constraint $h(\mathbf{x}) = 0$, $h : \Re^n \to \Re$ and $\mathbf{x}_0$ also be a reqular point (ie $\nabla h(\mathbf{x_0}^*) \neq 0$), then there exists a scalar λ^* such that

$\nabla f(\mathbf{x_0}^*) + \lambda^* \nabla h(\mathbf{x_0}^*) = \mathbf{0}$

that is $\nabla f(\mathbf{x_0}^*)$ and $\nabla h(\mathbf{x_0}^*)$ are parallel.

Note that the theorem above provides a first-order necessary condition for a point to be a local minimizer of $f(\cdot)$ subject to an equality constraint.

1.1.1 Definition (as reminder)

1. The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent if and only if the equation $\lambda_1 \mathbf{u}_1 + \ldots + \lambda_n \mathbf{u}_n = \mathbf{0}$ has as solution *only* the all-zero vector $(\lambda_1, \ldots, \lambda_n) = (0, \ldots, 0)$.

2. If the equation above has more solutions (essentially

non-zero) other than the all-zero one, then vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are called linearly dependent.

Note: One vector $\mathbf{u}$ by itself is linearly dependent or independent? If $\mathbf{u} \neq \mathbf{0}$ then $\lambda \mathbf{u} = \mathbf{0} \Rightarrow \lambda = 0 \Rightarrow \mathbf{u}$: linearly independent. But if $\mathbf{u} = \mathbf{0}$ then the equation $\lambda \mathbf{u} = 0$ has several (infinite) solutions. So, $\mathbf{u} = 0$ is linearly dependent.

Example:

Consider the surface

$\mathcal{S} = \left\{(x_1, x_2) : h(x_1, x_2) = x_1^2 = 0\right\} = \{(0, x_2), x_2 \in \Re\}$.

We have

$$\nabla h(x_1, x_2) = \begin{bmatrix} 2x_1 & 0 \end{bmatrix}$$

$$\begin{cases} \text{if } x_1 \neq 0 \text{ then } \nabla h \text{ is linearly independent} \\ \text{if } x_1 = 0 \text{ then } \nabla h \text{ is linearly dependent} \end{cases}$$

and

$$\mathcal{M} = \left\{(y_1, y_2) : \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{0}\right\}$$

In this example we cannot define the tangent plane at non-regular points. From now on, unless otherwise stated, we will consider surfaces $\mathcal{S} = \{\mathbf{x} : h_i(\mathbf{x}) = 0,\ i = 1, \ldots., m\}$ that have all their points regular.

That is the reason why in the Lagrange theorem point $\mathbf{x_0}$ must be regular.

1.1.2 Example

An Example on Lagrange theorem is shown in Figure 1 (where note that the curve $h = 0$ corresponding to the constraints should be intersecting with line $f = f_2$ at a point $\mathbf{x}_0$ so that the gradients of f and h are at the same point).

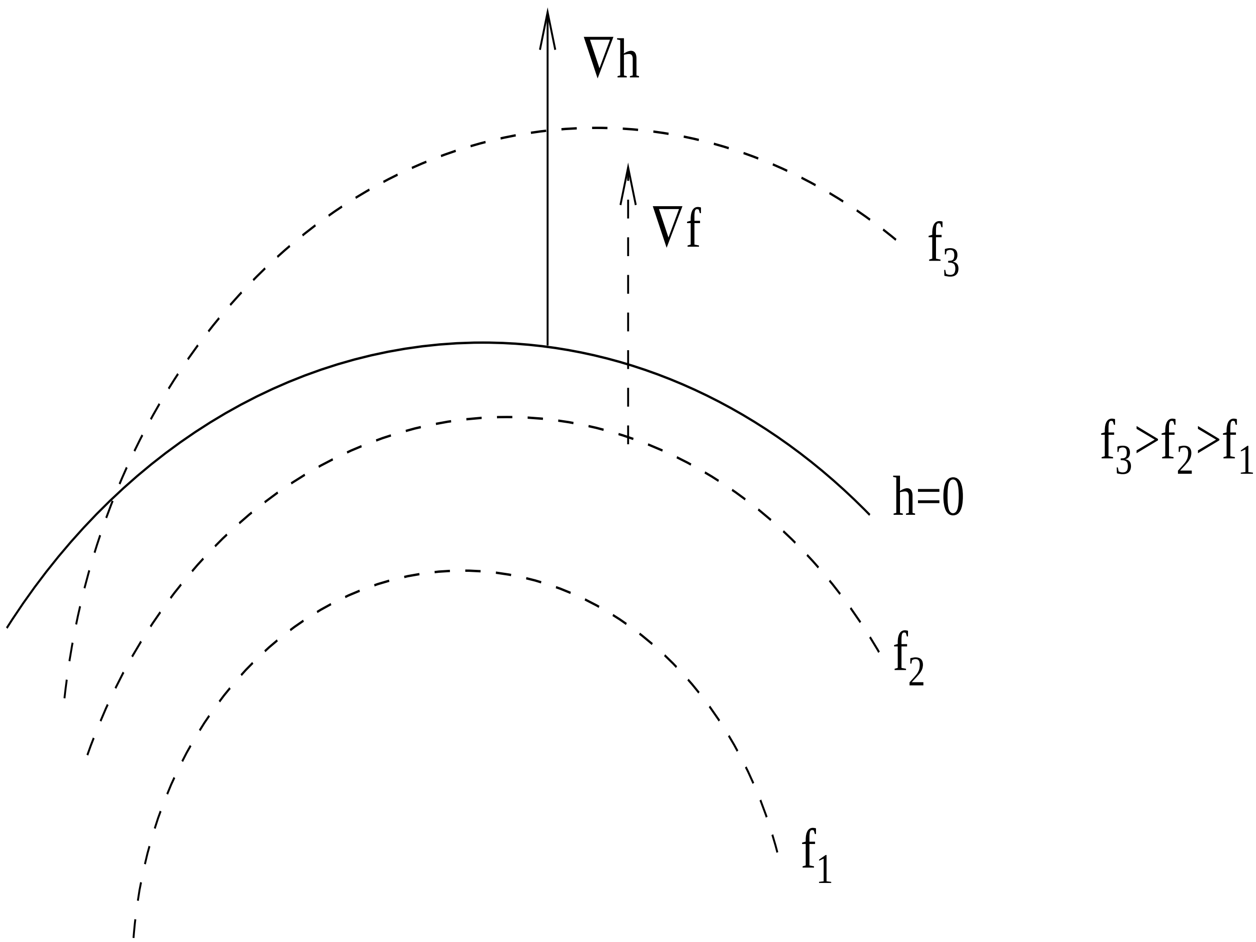

Figure 1: Example for the theorem of Lagrange.

1.2 Lagrange Theorem for $m > 1$ constraints

The Lagrance Theorem for $m > 1$ constraints becomes as follows:

Let the point $\mathbf{x_0}$ be a local minimizer of $f : \Re^n \to \Re$ subject to the constraints $h_1(\mathbf{x}) = 0, \ldots, h_m(\mathbf{x}) = 0$. Assume that $\mathbf{x_0}$ is a regular point. Then, there exists a real vector $\lambda^* \in \Re^m : \nabla f(\mathbf{x_0}) + \sum_{i=1}^{m} \lambda_i{}^* \ \nabla h(\mathbf{x_0}) = 0$.

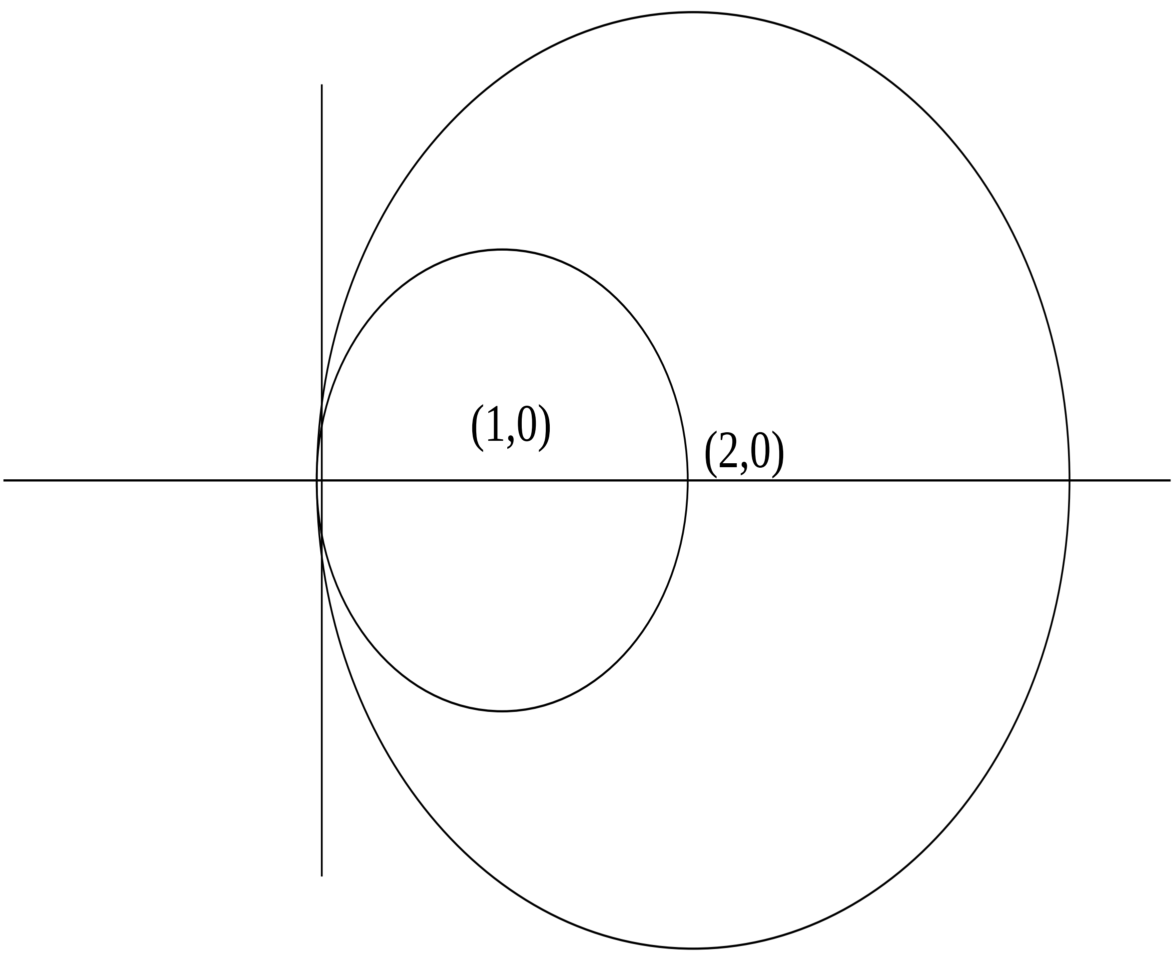

Figure 2: Lagrange theorem..

Thus, at an optimal point (if this is optimal), the gradient of the objective function can be written as linear combination of the gradients of the constraints.

1.2.1 Example

Consider the following problem which is depicted in figure 2:

$$\begin{array}{lll} \min & f(x_1, x_2) = x_1 + x_2 & \\ \text{s.t} & (x_1 - 1)^2 + x_2^2 - 1 = 0, & (h_1(x_1, x_2) = 0) \\ & (x_1 - 2)^2 + x_2^2 - 4 = 0, & (h_2(x_1, x_2) = 0) \end{array}$$

We have:

$$\nabla h_1(x_1, x_2) = (2(x_1 - 1), 2x_2)$$
$$\nabla h_2(x_1, x_2) = (2(x_1 - 2), 2x_2)$$

The surface $\mathcal{S}$ is the point $(0, 0)$. So we have

$$\nabla h_1(0, 0) = (-2, 0)$$
$$\nabla h_2(0, 0) = (-4, 0).$$

Now we try to confirm Lagrange theorem:

$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla h(\mathbf{x}^*) = 0 \Rightarrow$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} -2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -4 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow$$

$1 - 2\lambda_1 - 4\lambda_2 = 0 \Rightarrow$

$1 = 0$

So, because of $(0, 0)$ is not a regular point, we can't apply Langrange theorem here! The condition cannot hold for any λ_1, λ_2, namely the gradient of the objective function cannot be expressed as a linear combination of the constraints.

1.3 Lagrangian function

Recall again the form of the problem we are considering here:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t} & h_1(\mathbf{x}) = 0 \\ & \vdots \\ & h_m(\mathbf{x}) = 0 \end{array}$$

Define the *Lagrangian function* at point $\mathbf{x}$ as:

$$\begin{aligned} L(\mathbf{x}, \lambda_1, \ldots, \lambda_m) &= f(\mathbf{x}) + \lambda_1 h_1(\mathbf{x}) + \ldots + \lambda_m h_m(\mathbf{x}) \\ &= f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}) \qquad (5) \end{aligned}$$

The Hessian matrix of the Lagrangian at point $\mathbf{x}$, $\Lambda(\mathbf{x}, \lambda_1, \ldots, \lambda_m)$ is defined as

$$\Lambda(\mathbf{x}, \lambda_1, \ldots, \lambda_m) = F(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i H_i(\mathbf{x}), \qquad (6)$$

where $F(\mathbf{x})$ is the Hessian matrix of the objective function f at point $\mathbf{x}$, given by

$$F(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_m} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

where the partial derivatives are evaluated at point $\mathbf{x}$ and $H_i(\mathbf{x})$ is the Hessian matrix of h_i at point $\mathbf{x}, i = 1, .., m$.

2 Necessary and Sufficient Conditions

We now state the necessary conditions of second-order for the existence of local minimum. Then, we state second-order sufficient conditions for existence of local minimum.

First order necessary conditions for existence of local minimum: Let point $\mathbf{x}^*$ be a local minimizer of $f : \Re^n \rightarrow \Re$ subject to the constraints $h_1(\mathbf{x}) = 0, ..., h_m(\mathbf{x}) = 0$. Suppose $\mathbf{x}^*$ is a regular point. Then, there exists vector $\lambda^* \in \Re^m$ such that:

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \ \nabla h_i(\mathbf{x}^*) = \mathbf{0} \tag{7}$$

Second order necessary conditions for existence of local minimum: Let $\mathcal{M} = \{\mathbf{y} : \nabla h_i(\mathbf{x}^*)\mathbf{y} = 0, i = 1, ..., m\}$ be the tangent plane at point $\mathbf{x}^*$ of the surface defined by the equality constraints. Then, matrix

$$\Lambda\left(\mathbf{x}^*, \lambda^*\right) = F(\mathbf{x}^*) + \lambda_1{}^* \ H_1(\mathbf{x}^*) + ... + \lambda_m{}^* \ H_m(\mathbf{x}^*) \tag{8}$$

is positive-semidefinite on $\mathcal{M}$. That is,

$$\mathbf{y}^T \Lambda\left(\mathbf{x}^*, \lambda^*\right) \mathbf{y} \geq 0, \ \forall \mathbf{y} \in \mathcal{M}, \ \ \mathbf{y} \neq 0 \tag{9}$$

Sufficient conditions for existence of local minimum:

If there exists a real vector $\boldsymbol{\lambda} \in \Re^m$ which, at a point $\mathbf{x}^*$ satisfies

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \ \nabla h_i(\mathbf{x}^*) = 0 \tag{10}$$

and if matrix

$$\Lambda\left(\mathbf{x}^*, \lambda^*\right) = F(\mathbf{x}^*) + \lambda_1{}^* \ H_1(\mathbf{x}^*) + ... + \lambda_m{}^* \ H_m(\mathbf{x}^*) \tag{11}$$

is positive definite on $\mathcal{M}$
(where $\mathcal{M} = \left\{\mathbf{y} : \nabla h_i(\mathbf{x}^*)^T \mathbf{y} = 0, i = 1, ..., m\right\}$) that is, $\mathbf{y}^T \Lambda\left(\mathbf{x}^*, \lambda^*\right) \mathbf{y} > 0, \forall \mathbf{y} \in \mathcal{M}, \mathbf{y} \neq 0$,
then $\mathbf{x}^*$ is a local minimizer of $f : \Re^n \to \Re$ subject to the constraints $h_i(\mathbf{x}) = 0, i = 1, .., m$.

Note that, unless otherwise stated in the cases we will consider, we will assume that $\mathcal{M} = \Re^n$.

2.1 Sufficient conditions minimum for *convex* functions

Assume that the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is a *convex* function and the constraint functions $h_i(\mathbf{x})$, with $h_i : \mathbb{R}^n \to \mathbb{R}$ are also convex functions which define the surface of constraints $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n, h_i(\mathbf{x}) = 0, i = 1, \ldots, m\}$. If there exists $\mathbf{x}^* \in \mathcal{S}$ and $\lambda^* \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \nabla h_i(\mathbf{x}^*) = 0$$

then the $\mathbf{x}^*$ is global minimum of f.

Note 1: In the case of convex functions, the sufficient conditions for existence of global minimum are obtained only by equating the partial derivatives of the Lagrangian to zero. The Hessian matrices of functions f, h_i are positive-definite and the second condition is not needed. In the case of convex functions, the local minimum is global minimum.

Note 2: The factors $\lambda_i, i = 1, \ldots, m$ are called *Lagrange multipliers* regardless if functions are convex or not.

2.2 Local and global maximum of NLP problems

For the maximization problem

$$\max f(\mathbf{x})$$
$$\text{s.t. } h_1(\mathbf{x}) = 0\,, \ldots, h_m(\mathbf{x}) = 0$$

the second-order sufficient conditions for existence of local maximum are as follows: If there exist $\mathbf{x}^*, \lambda^*$ such that the gradient of the Lagrangian function is zero,

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \nabla h(\mathbf{x}^*) = \mathbf{0}$$

and the Hessian matrix of the Lagrangian

$$\Lambda(\mathbf{x}^*, \lambda^*) < 0,$$

i.e, the matrix is negative-definite, then $\mathbf{x}^*$ is local maximum of f under the constraints $h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m$.

Note: If $f(\mathbf{x})$ is a **concave** function and the constraint functions $h_i(\mathbf{x}),\ i = 1, \ldots, m$ are also concave, then $\mathbf{x}^*$ is a global maximum of f in the constrained maximization problem.

2.3 Examples

2.3.1 Example 1

Consider the minimization problem,

$$\min \tfrac{1}{2}\mathbf{x}^T Q\mathbf{x}$$
$$\text{subject to } A\mathbf{x} = \mathbf{b}$$

where $Q > 0$ is symmetric, positive definite matrix, $A \in \mathbb{R}^{m\times n}, m < n$ and $b \in \mathbb{R}^m$.

From equation $A\mathbf{x} = b$ we get a Lagrangian multiplier vector $\boldsymbol{\lambda}$. The Lagrangian function is:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{b} - A\mathbf{x})$$

The objective function $\frac{1}{2}\mathbf{x}^T Q\mathbf{x}$ is convex because $\nabla f(\mathbf{x}) = Q\mathbf{x}$ and the Hessian is $F(\mathbf{x}) = Q > 0$. The Lagrange condition for existence of minimum is:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0} \quad \Rightarrow Q\mathbf{x} - A^T \boldsymbol{\lambda} = \mathbf{0}$$

and the optimal solution satisfies:

$$\mathbf{x}^* = Q^{-1} A^T \boldsymbol{\lambda}$$

To find $\boldsymbol{\lambda}$, we use the fact that $\mathbf{x}^*$ is a feasible point, so it satisfies the constraints. So:

$$A\mathbf{x} = \mathbf{b} \quad \Rightarrow$$

$$AQ^{-1}A\boldsymbol{\lambda} = \mathbf{b} \quad \Rightarrow$$

$$\boldsymbol{\lambda} = (AQ^{-1}A)^{-1}\mathbf{b}$$

Thus,

$$\mathbf{x}^* = Q^{-1} A^T (AQ^{-1}A^T)^{-1} \mathbf{b} \tag{12}$$

is the global minimum of our problem.

2.3.2 Example 2

Consider the problem

$$\max f(\mathbf{x}) = x_1 x_2 + x_2 x_3 + x_1 x_3$$
$$\text{subject to: } x_1 + x_2 + x_3 = 3.$$

Solution: The Lagrangian function is:

$$L(x_1, x_2, x_3, \lambda) = x_1x_2+x_2x_3+x_1x_3+\lambda(x_1+x_2+x_3-3)$$

We equate the partial derivatives $\frac{\partial L}{\partial x_i} = 0, i = 1, 2, 3$, and we get the equations below:

$$x_2 + x_3 + \lambda = 0 \tag{13}$$

$$x_1 + x_3 + \lambda = 0 \tag{14}$$

$$x_1 + x_2 + \lambda = 0 \tag{15}$$

and we also have the constraint:

$$x_1 + x_2 + x_3 = 3 \tag{16}$$

Solving the 4×4 system of equations, we have:

$$x_1^* = 1, \quad x_2^* = 1, \quad x_3^* = 1, \quad \lambda = -2$$

For the objective function $f(\mathbf{x})$, we get the gradient vector:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{pmatrix}$$

The Hessian matrix is:

$$F(\mathbf{x}) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

and does not depend on $\mathbf{x}$. The Hessian matrix for the constraint $h(x) = x_1 + x_2 + x_3 - 3$ is:

$$H(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Thus,

$$\Lambda(\mathbf{x}^*, \lambda) = F(\mathbf{x}^*) + \lambda H(\mathbf{x}^*) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

since $H(\mathbf{x}^*) = 0$. Now, we don't know whether the Hessian matrix is positive- or negative- definite in $\mathbb{R}^3$. As a result,we cannot claim if $\forall \mathbf{y} \in \mathbb{R}^3 : \mathbf{y}^T \Lambda \mathbf{y}$ is a positive or negative quantity. But, we take into account the *precise* formulation of the sufficient condition, which states that the Hessian should be positive-definite (negative-definite) on the tangent plane to the surface defined by the constraints, so as to have local minimum (local maximum) in the problem. The tangent plane to the surface defined by the (one and only) constraint in the problem,

$$h(\mathbf{x}) = 0 \quad \Rightarrow \quad x_1 + x_2 + x_3 - 3 = 0$$

is

$$
\begin{aligned}
\mathcal{M} &= \{\begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} : \nabla h(\mathbf{x})^T \mathbf{y} = 0\} \\
&= \{\mathbf{y} : \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = 0\} \Rightarrow
\end{aligned}
$$

and finally

$$
\mathcal{M} = \{ \quad (y_1, y_2, y_3) : y_1 + y_2 + y_3 = 0\} \qquad (17)
$$

We examine if the Hessian matrix is positive- or negative-definite at the tangent plane $\mathcal{M}$. We have

$$
\begin{aligned}
\mathbf{y}^T \Lambda \mathbf{y} &= \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\
&= y_1(y_2 + y_3) + y_2(y_1 + y_3) + y_3(y_1 + y_2)
\end{aligned}
$$

We have:

$$
y_2 + y_3 = -y_1, \quad y_1 + y_3 = -y_2, \quad y_1 + y_2 = -y_3
$$

and thus $\mathbf{y}^T \Lambda \mathbf{y} = -y_1^2 - y_2^2 - y_3^2 \leq 0$. Thus, the Hessian matrix is negative-definite and $\mathbf{x}^* = (1, 1, 1)$ that was found above is local maximum to the problem.

2.3.3 Example 3

Consider the problem

$$\max \frac{\mathbf{x}^T Q \mathbf{x}}{\mathbf{x}^T P \mathbf{x}} \tag{18}$$

with matrix $Q = Q^T \geq 0$ and $P = P^T > 0$ (Q, P are symmetric and positive-definite matrices).

In the problem above, if $\mathbf{x}$ is an optimal solution, then all multiples $a\mathbf{x}$, $a \neq 0, a \in \mathbb{R}$ are optimal solutions too. In order to avoid the multiplicity of solutions, we set $\mathbf{x}^T P \mathbf{x} = 1$ and we get the constrained problem:

$$\max \mathbf{x}^T Q \mathbf{x}$$
$$\text{subject to: } \mathbf{x}^T P \mathbf{x} = 1$$

The Lagrangian function is:

$$L(\mathbf{x}, \lambda) = \mathbf{x}^T Q \mathbf{x} + \lambda(1 - \mathbf{x}^T P \mathbf{x}) \tag{19}$$

and by the condition $\nabla L(\mathbf{x}, \lambda) = \mathbf{0}$
$\Rightarrow 2Q\mathbf{x} - 2\lambda P\mathbf{x} = \mathbf{0} \Rightarrow Q\mathbf{x} = \lambda P\mathbf{x} \Rightarrow P^{-1}Q\mathbf{x} = \lambda\mathbf{x}$.

From the above, we observe that if $\mathbf{x}$ is a solution and maximizes $\mathbf{x}^T Q \mathbf{x}$ then it is an eigenvector which corresponds to some eigenvalue λ of matrix $P^{-1}Q$.
Thus, suppose that $\mathbf{x}^*$ is optimal solution, then we have:
$\mathbf{x}^{*T} P \mathbf{x}^* = 1$

and then
$P^{-1}Q\mathbf{x}^* = \lambda^*\mathbf{x}^* \;\Rightarrow\; PP^{-1}Q\mathbf{x}^* = \lambda^* P\mathbf{x}^* \;\Rightarrow$
$Q\mathbf{x}^* = \lambda^* P\mathbf{x}^* \;\Rightarrow\; \mathbf{x}^{*T}Q\mathbf{x}^* = \lambda^*\mathbf{x}^{*T}P\mathbf{x}^* \quad\Rightarrow$
$\lambda^* = \mathbf{x}^{*T}Q\mathbf{x}^*$
where λ^* is the Lagrange multiplier at the optimal solution.

Note that λ^* must be one of the n eigenvalues of $P^{-1}Q$, which are $\lambda_1 < \lambda_2 <$ ... $< \lambda_n)$. In particular, λ^* is the maximum eigenvalue of matrix $P^{-1}Q$ and the optimal solution $\mathbf{x}^*$ is the eigenvector which corresponds to the maximum eigenvalue λ^* of $P^{-1}Q$.

2.3.4 Example 4

Solve the problem

$$\min \mathbf{c}^T\mathbf{x}, \tag{20}$$

subject to:

$$\sum_{i=1}^{n} x_i = 0 \text{ and } \sum_{i=1}^{n} x_i^2 = 1 \tag{21}$$

Solution:

$$x_i^* = \frac{c_i + \lambda^*}{\mu^*}, \tag{22}$$

with

$$\lambda^* = -\frac{1}{n}\sum_{i=1}^{n} c_i, \text{ and } \mu^* = \pm\frac{1}{2}n^2\sqrt{\frac{1}{n}\sum_{i=1}^{n} c_i^2 - \frac{1}{n}(-n\lambda^*)^2} \tag{23}$$

3 Interpretation of Lagrange multipliers

Consider again the NLP problem with one equality constraint,

$$\min f(\mathbf{x})$$
$$\text{s.t. } h(\mathbf{x}) = 0$$

Let λ be the Lagrange multiplier corresponding to the one equality constraint. Assume a small variation c in the right-hand side of the constraint. Note that, as in the LP case, the constraint often specifies requirements in resources. Let $\mathbf{x}^*(0)$ be the optimal solution to the problem by having the constraint $h(\mathbf{x}) = 0$ and the value of the objective function is $f(\mathbf{x}^*(0))$. Let $\mathbf{x}^*(c)$ be the optimal solution by having the constraint $h(\mathbf{x}) = c$, $c \in \mathbb{R}$, and let the corresponding value of the objective function be $f(\mathbf{x}^*(c))$.

We calculate the rate of change of the value of the objective function for small variations of the constraints,

$$\begin{aligned}
\frac{df(\mathbf{x}(c))}{dc} &= \frac{\partial f(\mathbf{x}(c))}{\partial x_1}\frac{dx_1(c)}{dc} + \ldots + \frac{\partial f(\mathbf{x}(c))}{\partial x_n}\frac{dx_n(c)}{dc} \\
&= \nabla f(\mathbf{x}(c)^T \mathbf{x}'(c) \qquad (24)
\end{aligned}$$

where $\mathbf{x}'(c)$ is the vector of derivatives $(dx_1(c)/dc, \ldots, dx_n(c)/dc)$.
From the first-order conditions we have

$$\nabla f(\mathbf{x}(c)) + \lambda \nabla h(\mathbf{x}(c)) = \mathbf{0} \Rightarrow$$

$$\frac{\partial f(\mathbf{x}(c))}{\partial x_i} + \lambda \nabla \frac{\partial h(\mathbf{x}(c))}{\partial x_i} = 0 \text{ for } i = 1, \ldots, m \Rightarrow$$

$$\frac{\partial f(\mathbf{x}(c))}{\partial x_i} = -\lambda \frac{\partial h(\mathbf{x}(c))}{\partial x_i} \text{ for } i = 1, \ldots, m.$$

We substitute in the equation above and we have:

$$\frac{df(\mathbf{x}(c))}{dc} =$$

$$= -\lambda \left(\frac{\partial h(\mathbf{x}(c))}{\partial x_1}\frac{dx_1(c)}{dc} + \ldots + \frac{\partial h(\mathbf{x}(c))}{\partial x_n}\frac{dx_n(c)}{dc} \right)$$

$$\Rightarrow \frac{df(\mathbf{x}(c))}{dc} = -\lambda \qquad (25)$$

since if we differentiate both sides of the constraint $h(\mathbf{x}) = c$ with respect to c, we get that

$$\frac{\partial h(\mathbf{x}(c))}{\partial x_1}\frac{dx_1(c)}{dc} + \ldots + \frac{\partial h(\mathbf{x}(c))}{\partial x_n}\frac{dx_n(c)}{dc} = 1 \quad (26)$$

Therefore, λ is interpreted as the rate of change in the value of the objective function per unit of change in the resources. Thus it represents the *price* of the unit of constraint requirement.

Note 1: If a small variation of a contraint causes decrease in the objective's value,then λ is positive for this specific costraint. Respectively, if it causes increase in the objective's value, then λ is negative for this specific costraint.

Note 2: Remember that we did the interpretation of Lagrange multipliers in LP problem in the same spirit, where we had prove that $\frac{\Delta z}{\Delta b} = \lambda$.

3.1 Sensitivity analysis for $m > 1$ constraints

Consider the NLP problem with more than one constraints:

$$\min f(\mathbf{x})$$
$$\text{s.t. } h_1(\mathbf{x}) = 0, \ldots, h_m(\mathbf{x}) = 0.$$

Note that all constraints can be collectively described by vector $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. Let $\boldsymbol{\lambda}$ be the vector of Lagrange multipliers associated with the constraints.

Now let $\mathbf{c} \in \mathbb{R}^m$ denote a vector of small changes in the right-hand sides of the constraints. Similarly with the case of one constraint, we can define $\mathbf{x}(\mathbf{0})$ the optimal solution for constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and $\mathbf{x}(\mathbf{c})$ the optimal solution for constraints $\mathbf{h}(\mathbf{x}) = \mathbf{c}$. By following the methodology for the case of one constraint, we can prove:

$$\nabla_{\mathbf{c}} f(\mathbf{x}(\mathbf{c})) = -\boldsymbol{\lambda} \Rightarrow \left(\begin{array}{c} \frac{\partial f(\mathbf{x}(\mathbf{c}))}{\partial c_1} \\ \vdots \\ \frac{\partial f(\mathbf{x}(c))}{\partial c_n} \end{array} \right) = \left(\begin{array}{c} -\lambda_1 \\ \vdots \\ -\lambda_n \end{array} \right)$$

where

$$\lambda_i = \frac{\partial f(\mathbf{x}(\mathbf{c}))}{\partial c_i} \tag{27}$$

is again the price per unit of the resource i, or equivalently the rate of the of the value of the objective function with regard to small changes in the resource (constraint) i.

4 Beamforming

4.1 Starting from an example

We continue on the topic of Non-linear programming problems with equality constraints and we will solve the problem:

$$\min \mathbf{x}^T A \mathbf{x}$$

$$\text{s.t. } \mathbf{c}^T\mathbf{x} = 1$$

where $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{c} = (c_1, \ldots, c_n)$ and A is a matrix of dimension $n \times n$. Let λ be the Lagrange multiplier associated with the constraint of the problem. Thus the Lagrangian is,

$$L(\mathbf{x}, \lambda) = \mathbf{x}^T A\mathbf{x} + \lambda(\mathbf{c}^T\mathbf{x} - 1) \tag{28}$$

We use the conditions $\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = 0 \Leftrightarrow \frac{\partial L}{\partial x_i} = 0$ and we have:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = 0 \Rightarrow 2A\mathbf{x} + \lambda\mathbf{c} = 0 \Rightarrow \mathbf{x}^* = -\frac{1}{2}\lambda A^{-1}\mathbf{c}.$$

Now, we use the constraint $\mathbf{c}^T\mathbf{x} = 1$ to find the value of the Lagrange multiplier λ.

$$\mathbf{c}^T\mathbf{x} = 1 \Leftrightarrow -\frac{1}{2}\lambda\mathbf{c}^T A^{-1}\mathbf{c} = 1 \Leftrightarrow \lambda = \frac{-2}{\mathbf{c}^T A^{-1}\mathbf{c}}$$

Thus, the optimal soluton is

$$\mathbf{x}^* = \frac{A^{-1}\mathbf{c}}{\mathbf{c}^T A^{-1}\mathbf{c}} \tag{29}$$

As an application of this optimization problem, we will study the fundamental problem that arises in the case of beam-forming.

4.2 Beamforming Basics

In the case that we do not have a single omni-directional antenna but an array of omni-directional antennas, we can adapt

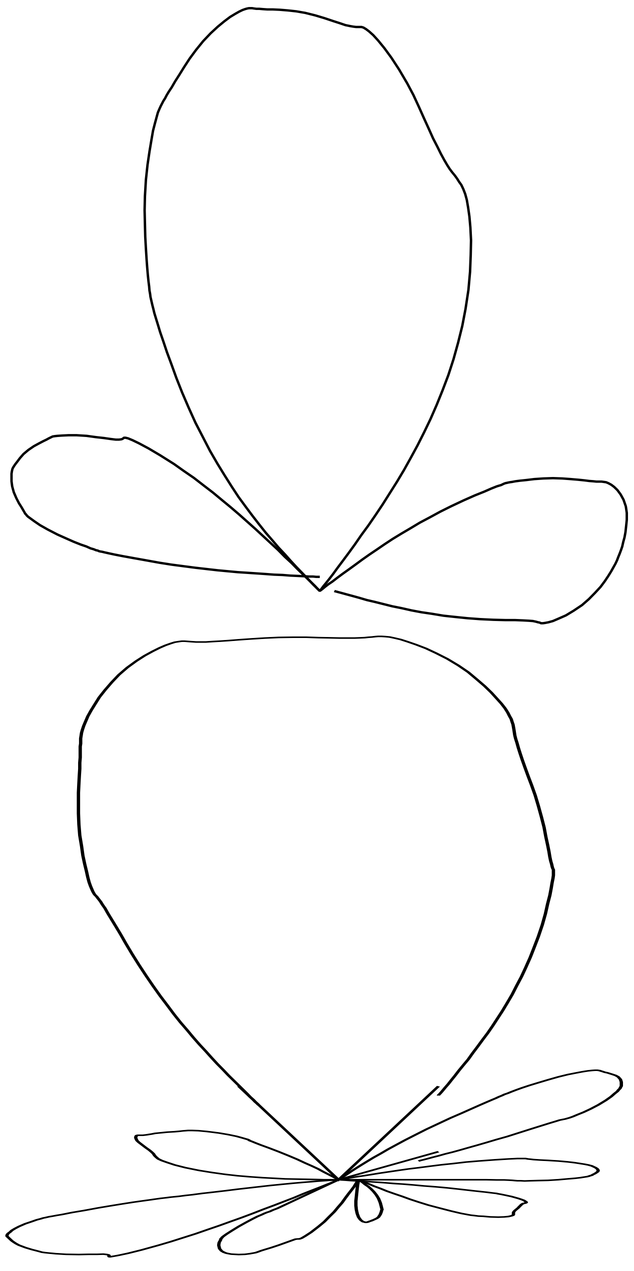

Figure 3: Different shapes of the antenna array radiation diagram as a result of controlling the electric current phases and amplitudes.

the radiation diagram by changing the amplitudes and phases of the alternate electric currents that feed the antenna. Thus, for example, for an antenna array of M elements, the radiation diagram (namely the width and length of the main lobe and the angle of the lobe) is a function of of the complex numbers $\{I_i e^{j\phi_i}\}_{i=1}^{M}$, where I_i is the amplitude and ϕ_i is the phase of the alternate current which stimulate the antenna element i. Thus, we can control the radiation diagram and dynamically change the shape and form and make diagrams like the ones depicted in figure 3 Note that in the radiation diagram, most of the transmission power is concentrated towards a given direction, that of the main lobe. There also exist several side lobes as well.

An antenna array of controllable radiation diagram is called *adaptive antenna array* or *smart antenna* and the control of the radiation diagram is called *beam-forming*. Clearly, the radiation diagram can be controlled either in order to transmit or to receive a signal. A smart antenna can adapt its radiation diagram according e.g. to the instantaneous position of the user. The following advantages exist for an adaptive antenna array:

1. Minimization of interference. Beam-forming can take place either in the reception or in transmission. We can shape the diagram in such a way that we can transmit or receive from a certain direction, that specified by the main lobe. For reception, the antenna array can receive signals only emitted from certain directions and atten-

uate signals emitted from other directions. The same holds for transmission.

2. Capacity increase. A transmitter can transmit at the same conventional channel (frequency or timeslot) to more than one users. Also, the same holds for the case of reception. In that case, a different radiation diagram is formed for each user. Clearly, there exists a M-fold increase in system capacity if the antenna array can serve at the same channel M users simultaneously (in the same frequency and time slot).

If we have M antenna elements in the antenna array, there can be $k \leq M$ radiation diagrams, one for each user. For simplicity assume that $M = 2$ here. The antenna can form at most two radiation diagrams and each diagram corresponds to a complex vector $\mathbf{w}_1 = (w_{11}, w_{12})$ where $w_{11} = I_1 e^{j\phi_1}$ and $w_{12} = I_2 e^{j\phi_2}$, and $\mathbf{w}_2 = (w_{21}, w_{22})$, defined similarly. The two vectors define the two radiation diagrams (basically the main lobes) and each radiation diagram can serve one user (if $\mathbf{w}_1$ and $\mathbf{w}_2$ are linearly independent, the antenna array can serve simultaneously both users). Each radiation diagram corresponds to a vector with dimension equal to the number of antenna elements.

We will now assume that the vectors are real numbers and we will not further worry about complex numbers. Still, the theory can be extended to cover the complex number case. We

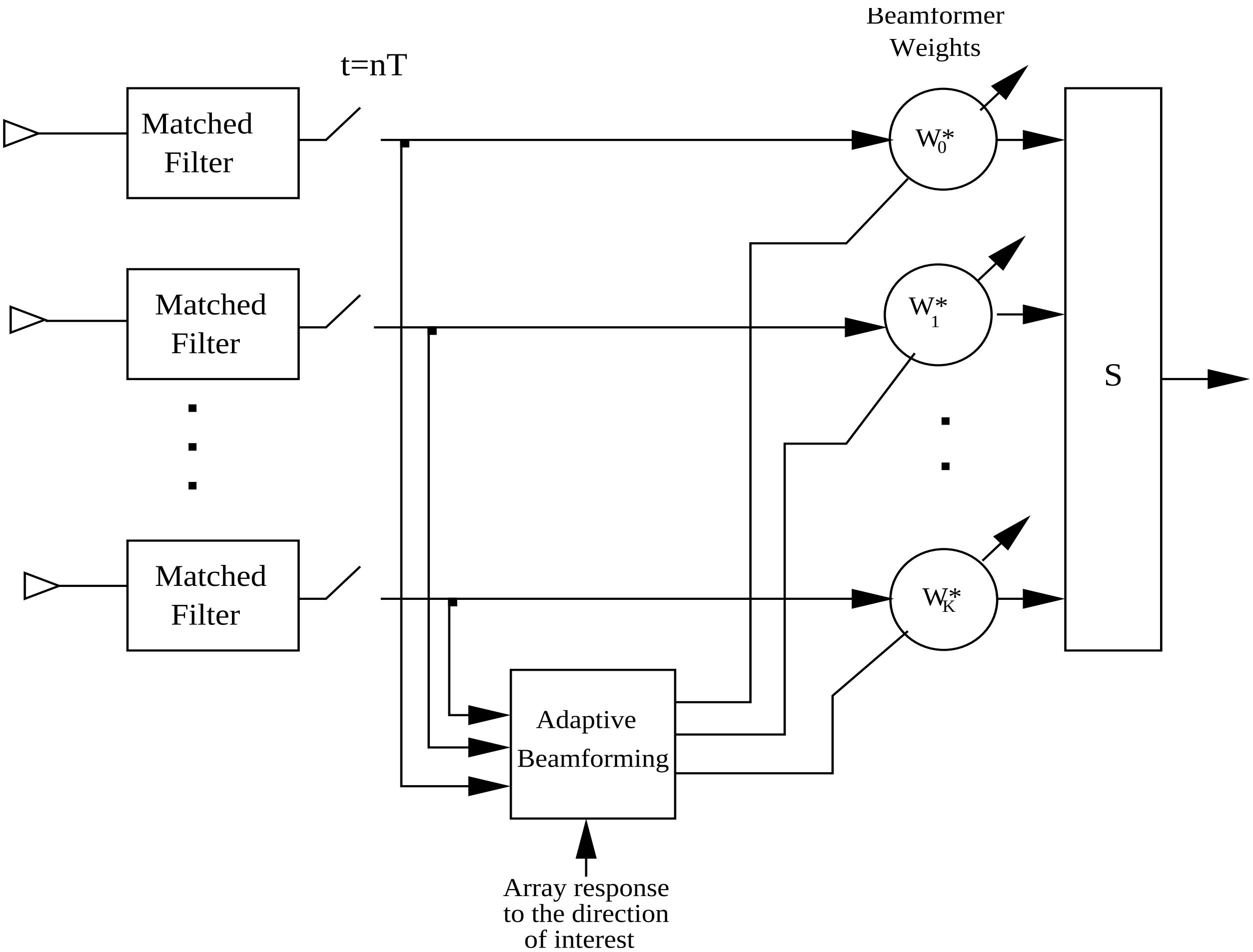

Figure 4: Block diagram of a receiver with an adaptive antenna array and a beamformer.

will also concentrate on the case where the signal of several users is received at a base station and the base station will attempt to discriminate the signal of only one user (the one of interest) by computing the beam-forming vector $\mathbf{w}$.

The main signal processing segment at the receiver is the adaptive beamformer (Figure 2). The adaptive beamformer finds vector $\mathbf{w}$. Its task is to find M numbers $(w_1, \ldots, w_M) = \mathbf{w}$. The signal y_i that has reached the i-th antenna is multiplied by w_i. Then we have to sum the above products to find the interior product $\mathbf{w}^T\mathbf{y} = w_1 y_1 + \ldots w_M y_M$ as the total outcome of *combining* all received signals at antennas by an appropriate number. That is, the output from each array element i is weighted by a weight w_i and added. The objective is to search for the w_i's such that the Signal-to-Noise Ratio (SNR) of the signal of the user of interest is maximized. The SNR is taken at the output, after the summation in figure 2.

We define the antenna array response vector to the direction of arrival θ as $\mathbf{v}(\theta) = (v_i(\theta), \ldots, v_M(\theta))$ that shows how each antenna receives a signal coming from an angle θ. The received signal (vector signal) at the M antennas at some frozen time t is

$$\mathbf{y}(t) = \sum_{j=1}^{K} \sqrt{P_j G_j} \sum_{\ell=1}^{L} a_j \mathbf{v}_j(\theta_\ell) s_j(t - \tau_j) + \mathbf{n}(t)$$

where:

K: number of users.

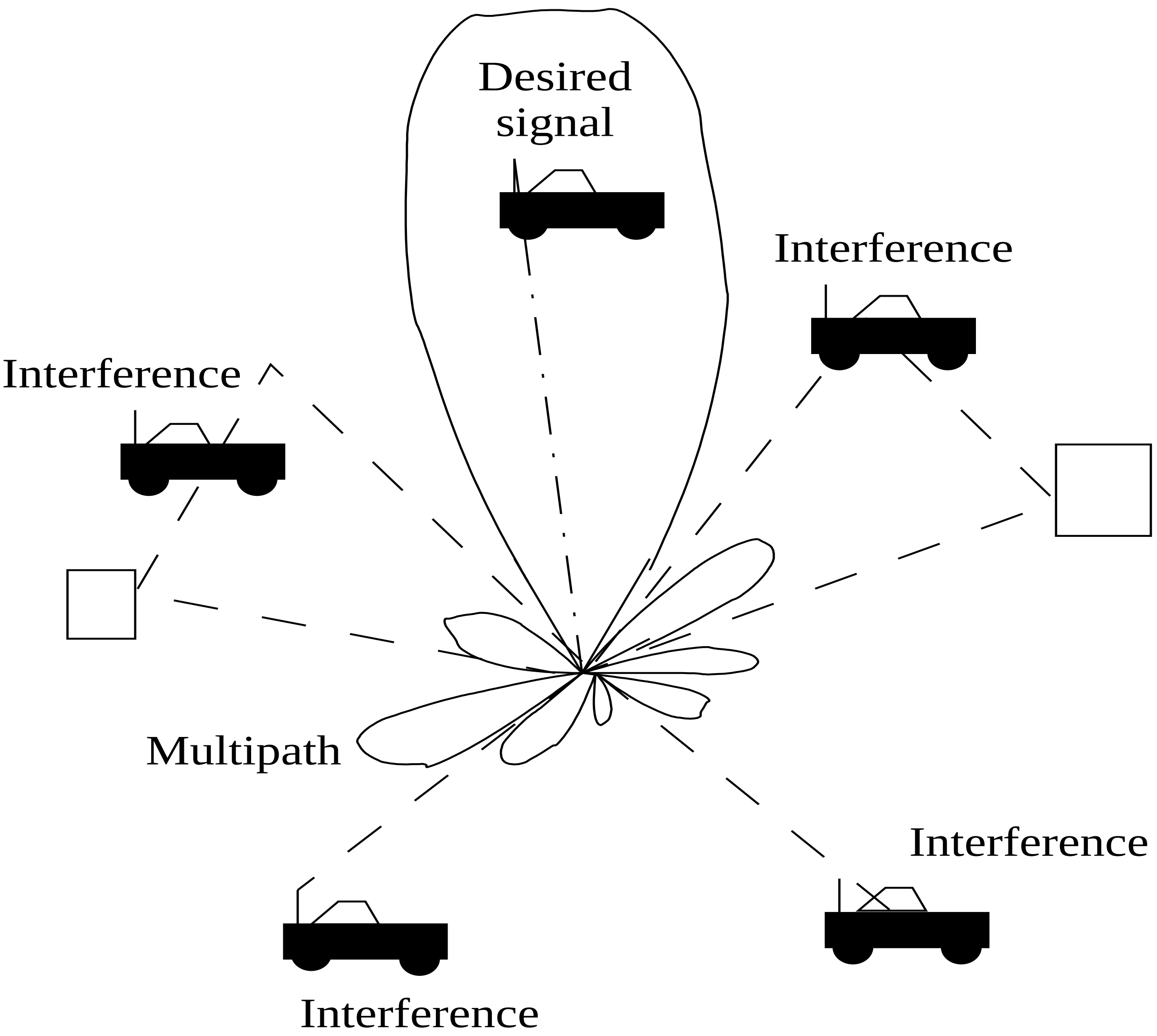

Figure 5: Example of an antenna array radiation diagram.

P_j: transmission power of the user j
G_j: path gain (denoting the distance loss) between user j and the BS. It is given by $G_j = \frac{1}{d_j^\gamma}$, where d_j is the distance from user j to the BS and γ is a constant that depends on the environment.
L: The number of paths of the multi-path (assume each user has its signal arriving through L paths).
Each of these (user j) paths has:
α_j^ℓ: Attenuation factor of path ℓ of user j because of shadowing (this is a random number, usually log-normally distributed).
θ_ℓ: angle of arrival of ℓ-th path.
$\mathbf{v}_j(\theta_\ell)$: Response vector of the antenna to a signal which is sent from user j and comes from path angle θ_ℓ.
s_j: the transmitted signal of user j.
τ_j: Signal delay for the signal of user j.
$\mathbf{n}(t)$: A vector that indicates noise at the receiver in each antenna.

Each user j can be completely specified by a vector called *spatial signature* of user j,

$$\mathbf{a}_j = \sum_{\ell=1}^{L} \alpha_j^\ell \mathbf{v}_j(\theta_\ell). \tag{30}$$

As we can see, the spatial signature depends on

- position of user i,

- number of paths,
- direction of arrival (DoA) of each path,
- shadowing coefficient of each path.

Thus, we have:

$$\mathbf{y}(t) = \sum_{j=1}^{k} \sqrt{P_j G_j}\mathbf{a}_j s_j(t-\tau_j) + \mathbf{n}(t)$$

Let the transmitted signal s_j by user j be represented as:

$$s_j(t) = \sum_{n} b_j(n) g(t-nT)$$

where,
$g(\cdot)$: pulse shaping filter function, specifying the shape of the pulse on which the bits will be carried.
T: The symbol time.
$\{b_j(n)\}, n = 1, \ldots,$: the sequence of bits.
At the receiver, we have the matched filter receiver (matched to the pulse shaping filter function of the transmitter) that is given by $g(t) = g^*(-t)$. The output of the matched filter is sampled at discrete times $t = nT$ (once in a symbol time) and we have the received discrete-time signal at the output of the matched filter as

$$\mathbf{y}(n) = \mathbf{y}(t) * g^*(-t)|_{t=nT}$$

Convolution of the received signal with the matched filter and sampling at $t = nT, n = 1, \ldots$ is equivalent to operation

$$\int_{(n-1)T}^{nT} \sum_n b(n) g(t - nT) g^* dt = b(n)$$

The receiver calculates the above integral in each symbol time interval. Thus, the signal from continuous-time becomes discrete-time:

$$\mathbf{y}(n) = \sum_{j=1}^{K} \sqrt{P_j G_j} \mathbf{a}_j b_j(n) + \mathbf{n}(n)$$

The expected power of the output signal after the beamforming and the multiplication with factors w_i is,

$$\mathbb{E}[e^2] = \mathbb{E}\left[\left| \mathbf{w}^T \mathbf{y} \right|^2 \right] = \mathbb{E}\left[\mathbf{w}^T \mathbf{y} \mathbf{y}^T \mathbf{w} \right] = \mathbf{w}^T \mathbb{E}[\underbrace{\mathbf{y}\mathbf{y}^T}_{A}] \mathbf{w}$$

Matrix A is of dimension $M \times M$ and each element $A_{ij} = \mathbb{E}[y_i y_j]$ shows the correlation between received signals at antennas i and j. We can see from the relation above that

$$A = \sum_{j=1}^{K} P_j G_j \mathbf{a}_j \mathbf{a}_j^T + \sigma^2 I$$

under the assumptions that user signals are zero-mean ($\mathbb{E}[|b_i(n)|] = 0$), different user signals are uncorrelated ($\mathbb{E}[b_i(n) b_j(n$

0, for $i \neq j$), user signals are unit-power ($\mathbb{E}[|b_i^2(n)|] = 1$). Also each random variable representing noise at each antenna is Gaussian with zero mean and variance σ^2, and the noise variables at different antennas are uncorrelated:

$$\mathbb{E}[n_i n_j] = \begin{cases} 0 & \text{,if } i \neq j \\ \sigma^2 & \text{, if } i = j \end{cases}$$

The base station receives data from all K users and needs to calculate the beam-forming vector $\mathbf{w}$ to distinguish the signal of a user i. We can write the matrix A as consisting of two parts, one concerning the user of interest i and another concerning all other users (which is essentially interference),

$$A = \underbrace{P_i G_i \mathbf{a}_i \mathbf{a}_i^T}_{A_i} + \underbrace{\sum_{j \neq i} P_j G_j \mathbf{a}_j \mathbf{a}_j^T + \sigma^2 I}_{A_{\text{int}}}$$

Thus, we have:

$$\mathbb{E}[\mathbf{w}^T A \mathbf{w}] = \mathbf{w}^T (P_i G_i \mathbf{a}_i \mathbf{a}_i^T + A_{\text{int}}) \mathbf{w} = \mathbf{w}^T A_{\text{int}} \mathbf{w} + P_i G_i \|\mathbf{w}^T$$

The expected signal power at the output comprises the signal power that originates from user i and the power that originates from all other users. Thus, we have for the signal-to-interference and noise ratio:

$$\text{SINR}_i = \frac{P_j G_j \|\mathbf{w}^T \mathbf{a}_i\|^2}{\mathbf{w}^T A_{\text{int}} \mathbf{w}},$$

The receiver wants to find the vector $\mathbf{w}$ to maximize SINR_i, so it faces the problem:

$$\max_{\mathbf{w}} \text{SINR}_i = \min_{\mathbf{w}} \frac{\mathbf{w}^T A_{\text{int}} \mathbf{w}}{P_j G_j \|\mathbf{w}^T \mathbf{a}_i\|^2}$$

This is equivalent to maintaining $\mathbf{w}^T \mathbf{a}_i = 1$ (fixed) at the direction of the user of interest and trying to minimize interference. Thus, the problem becomes:

$$\min_{\mathbf{w}} \mathbf{w}^T A_{\text{int}} \mathbf{w}$$

$$\text{subject to: } \mathbf{w}^T \mathbf{a}_i = 1$$

From the solution of the problem in the beginning of the lecture, we find that the optimal beam-forming vector $\mathbf{w}^*$ is:

$$\mathbf{w}^* = \frac{A_{\text{int}}^{-1} \mathbf{a}_i}{\mathbf{a}_i^T A_{\text{int}}^{-1} \mathbf{a}_i}$$

5 NLP problems with inequality constraints

In previous lectures, we studied non-linear programming problems with equality constraints. We will now generalize the theory to problems which also have inequality constraints. Namely, we will consider problems of the form:

$$\min f(x) \tag{31}$$

subject to:

$$h_i(\mathbf{x}) = 0,\ i = 1, ..., m,\ \text{ and } g_j(\mathbf{x}) \leq 0,\ j = 1, ..., p. \tag{32}$$

A point $\mathbf{x}_0$ is called *feasible* if it satisfies all constraints, namely it is $h_i(\mathbf{x}_0) = 0,\ i = 1, ..., m$ and $g_j(\mathbf{x}_0) \leq 0,\ j = 1, ..., p$.

An inequality constraint $g_j(\cdot)$ is called *active* at point $\mathbf{x_0}$ if it is satisfied with equality, namely it is $g_j(\mathbf{x}_0) = 0$, otherwise it is called *inactive*.

Let $\mathcal{J}(\mathbf{x}_0)$ be set of indices of inequality constraints that are active at point $\mathbf{x}_0$. A point $\mathbf{x}_0$ is called *regular point* of the constraints if the vectors $\nabla h_i(\mathbf{x}_0)$, for $i = 1, \ldots, m$ and $\nabla g_j(\mathbf{x}_0)$ for $j \in \mathcal{J}(\mathbf{x}_0)$ are linearly independent.

6 Necerrary and Sufficient Conditions

6.1 First-order Necessary Conditions for existence of local minimum

Define the Lagrangian function at point $\mathbf{x}$, as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j g_j(\mathbf{x}) \quad (33)$$

where λ_i, $i = 1, \ldots, m$ are the Lagrange multipliers corresponding to the equality constraints and μ_j, $j = 1, \ldots, p$ are the Karush-Kuhn-Tucker (KKT) multipliers corresponding to inequality constraints $g_j(\mathbf{x}) \leq 0$.

6.1.1 Karush-Kuhn-Tucker (KKT) theorem

Suppose that $\mathbf{x}^*$ is a regular point of constraints $h_i(\mathbf{x}) = 0$, $i = 1, ..., m$ and $g_j(\mathbf{x}) \leq 0$, $j = 1, ..., p$. If $\mathbf{x}^*$ is a local minimum of $f(\mathbf{x})$ subject to the constraints $h_i(\mathbf{x}) = 0$, $i = 1, ..., m$ and $g_j(\mathbf{x}) \leq 0$, $j = 1, ..., p$, then there exist vectors $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in {R^p}_+$ $(\mu^* \geq \mathbf{0})$ such that:

$$\nabla L(\mathbf{x}, \lambda^*, \mu^*) = \quad (34)$$

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} {\lambda_i}^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} {\mu_j}^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}$$

and

$$\sum_{i=1}^{p} {\mu_j}^* g_j(\mathbf{x}^*) = 0 \tag{35}$$

Note that in the second equation above, since ${\mu_j}^* \geq 0$ and $g(\mathbf{x}^*) \leq 0$, the fact that $\sum_{i=1}^{p} {\mu_j}^* g_j(\mathbf{x}^*) = 0$ means that each term of the summation, ${\mu_j}^* g_j(\mathbf{x}^*) = 0$.
This further means the following:

- if the j-th KKT multiplier is ${\mu_j}^* > 0$, then the corresponding constraint $g_j(\mathbf{x}^*) = 0$, i.e it is met with equality at the optimal solution $\mathbf{x}^*$.

- Also, if a constraint is satisfied with strict inequality at the optimal solution, i.e $g_j(\mathbf{x}^*) < 0$ then the corresponding KKT multiplier ${\mu_j}^* = 0$ and thus it doesn't affects the Langrancian function.

These conditions are reminiscent of **complementary slackness** ones we saw at Linear Programming.

6.1.2 Graphical interpretation of KKT theorem

Consider the problem of minimizing function $\mathbf{x}$ subject to three inequality constraints $g_1(\mathbf{x}) \leq 0$, $g_2(\mathbf{x}) \leq 0$ and $g_3(\mathbf{x}) \leq 0$.

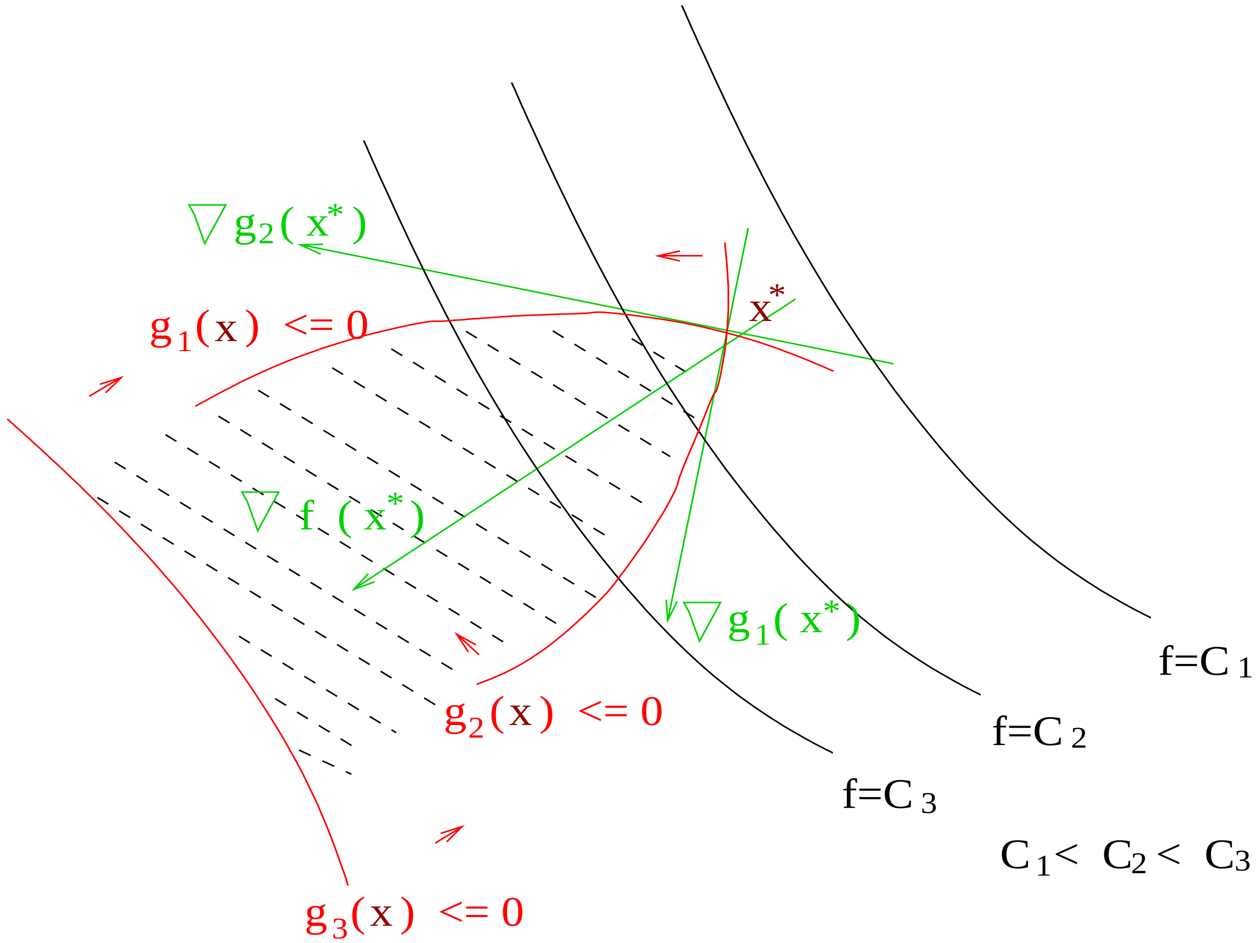

Figure 6: Graphical representation of the KKT theorem.

Each of the inequality constraints defines a subspace that is located on the one side of the curve $g_j(\mathbf{x}) = 0$.

Assume that at the optimal point, $\mathbf{x}^*$, we have that the active constraints are the first and the second, i.e it is $g_1(\mathbf{x}^*) = 0$ and $g_2(\mathbf{x}^*) = 0$, while the third one is inactive, $g_3(\mathbf{x}^*) < 0$. This case is depicted in figure 6.

The KKT theorem states that if $\mathbf{x}^*$ is local minimum of $f(\cdot)$, subject to the constraints $g_j(\cdot) \leq 0$, $j = 1, 2, 3$, then:

$$\nabla f(\mathbf{x}^*) + \mu_1 \nabla g_1(\mathbf{x}^*) + \mu_2 \nabla g_2(\mathbf{x}^*) + \mu_3 \nabla g_3(\mathbf{x}^*) = \mathbf{0} \tag{36}$$

Since $g_3(\mathbf{x}^*) < 0$ (inactive) $\longrightarrow \mu_3 = 0$ and we have

$$\nabla f(\mathbf{x}^*) = -\mu_1 \nabla g_1(\mathbf{x}^*) - \mu_2 \nabla g_2(\mathbf{x}^*) \tag{37}$$

that is, the gradient of f at point $\mathbf{x}^*$ is linear combination of the gradients of the active constraints at $\mathbf{x}^*$.

6.1.3 Example

Consider the problem

$$\min \; 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \tag{38}$$

subject to:

$$x_1^2 + x_2^2 \leq 5, \text{ and } 3x_1 + x_2 \leq 6 \tag{39}$$

The Lagrangian is:

$$\begin{aligned} L(x_1, x_2, \mu_1, \mu_2) \quad = \quad & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ & + \mu_1(x_1^2 + x_2^2 - 5) \\ & + \mu_2(3x_1 + x_2 - 6) \end{aligned}$$

From the KKT theorem, we have that if (x_1^*, x_2^*) is local minimum, then

$$\frac{\partial L(\cdot)}{\partial x_1}(x_1^*, x_2^*) = 0 \tag{40}$$

$$\frac{\partial L(\cdot)}{\partial x_2}(x_1^*, x_2^*) = 0. \tag{41}$$

So we get the equalities:

$$4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 = 0 \tag{42}$$

$$2x_1 + 2x_2 - 10 + 2\mu_1 x_2 + \mu_2 = 0 \tag{43}$$

Furthermore it is $\mu_1 \geq 0$, $\mu_2 \geq 0$ and also we have the *complementary slackness* conditions:

$$\mu_1(x_1^2 + x_2^2 - 5) = 0 \tag{44}$$

$$\mu_2(3x_1 + x_2 - 6) = 0 \tag{45}$$

Then if we want to find a point that satifies the necessary conditions of KKT theorem, we should start by taking cases and try various combinations of active constraints and check signs of the resulting KKT multipliers.

Assume that

$\mu_2 = 0 \rightsquigarrow 3x_1 + x_2 - 6 < 0$

and that

$\mu_1 > 0 \rightsquigarrow x_1^2 + x_2^2 - 5 = 0$

Then, for the conditions of the partial derivatives, we replace with $\mu_2 = 0$ and we have:

$$4x_1+2x_2-10+2\mu_1 x_1 = 0 \text{ and } 2x_1+2x_2-10+2\mu_1 x_2 = 0. \tag{46}$$

Also, since $\mu_1 > 0$, the first constraint holds with equality, $x_1^2 + x_2^2 - 5 = 0$.

We solve the system of equations and find $x_1^* = 1$, $x_2^* = 2$, $\mu_1^* = 1$ and this gives $3x_1^* + x_2^* - 6 = -1 < 0$ and thus the second constraint is satified.

Similarly, we can try other cases:

$\mu_1 > 0, \mu_2 > 0$

$\mu_1 = 0, \mu_2 > 0$

$\mu_1 = 0, \mu_2 = 0$.

6.2 Other forms of optimization problems

We saw that when we have the minimization problem

$$\min f(\mathbf{x}) \tag{47}$$

subject to:

$h_i(\mathbf{x}) = 0,\ \ i = 1, ..., m$ and $g_j(\mathbf{x}) \leq 0,\ \ j = 1, ..., p$

then we have the condition of the gradient of the Lagrangian being zero and also $\mu_j \geq 0$ for $j = 1, \ldots, p$.

What happens now if we have the problem:

$$\max f(\mathbf{x}) \tag{48}$$

subject to:

$h_i(\mathbf{x}) = 0,\ \ i = 1, ..., m$ and $g_j(\mathbf{x}) \leq 0,\ \ j = 1, ..., p$.

Write the objective as: $\max f(\mathbf{x}) = -\min f(\mathbf{x})$
subject to:
$h_i(\mathbf{x}) = 0,\ \ i = 1, ..., m$ and $g_j(\mathbf{x}) \leq 0,\ \ j = 1, ..., p$.

How will the KKT change? Let us apply the KKT theorem to the minimization problem of $-f(\mathbf{x})$.

If $\mathbf{x}^*$ is local maximum then

$$\nabla_{\mathbf{x}} L(\mathbf{x}^*, \lambda^*, \mu^*) = \tag{49}$$

$$\nabla(-f(\mathbf{x}^*)) + \sum_{i=1}^{m} {\lambda_i}^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} {\mu_j}^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}$$

and also:

$$\mu^* \geq \mathbf{0} \tag{50}$$

and

$$ {\mu_j}^* g_j(\mathbf{x}) = 0,\ \ \forall j \tag{51}$$

Multiplying with (-):

$$\nabla f(\mathbf{x}^*) - \sum_{i=1}^{m} {\lambda_i}^* \nabla h_i(\mathbf{x}^*) - \sum_{j=1}^{p} {\mu_j}^* \nabla g_j(\mathbf{x}^*) = \mathbf{0} \Rightarrow \tag{52}$$

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j{}^* \nabla g_j(\mathbf{x}^*)) = \mathbf{0}, \tag{53}$$

with $\mu_j{}^* \leq 0$.
Thus, if $\mathbf{x}^*$ is local maximum, then $\mathbf{x}^*$ satisfies $\nabla L(\mathbf{x}^*, \lambda^*, \mu^*) = \mathbf{0}$ and now it should be $\mu^* \leq \mathbf{0}$.
The condition $\mu_j{}^* g_j(\mathbf{x}^*) = 0$ for each $j = 1, \ldots, p$ should also hold.

Now assume we have the problem:

$$\min f(\mathbf{x})$$

subject to:
$h_i(\mathbf{x}) = 0,\ i = 1, \ldots, m$ and $g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, p$
i.e the inequalities are now reversed.
We multiply with -1, so as to bring the inequality in the usual form: $-g_j(\mathbf{x}) \leq 0$

Then, the KKT theorem is: If $\mathbf{x}^*$ local minimum of f, then:

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j{}^* (-g_j(\mathbf{x}^*)) = \mathbf{0} \Rightarrow \tag{54}$$

$$\nabla f(\mathbf{x}^*)+\sum_{i=1}^{m}\lambda_i{}^*\nabla h_i(\mathbf{x}^*)-\sum_{j=1}^{p}\mu_j{}^*\nabla g_j(\mathbf{x}^*)=\mathbf{0},\ \ \mu_j{}^*\geq 0 \tag{55}$$

or

$$\nabla f(\mathbf{x}^*)+\sum_{i=1}^{p}\lambda_i{}^*\nabla h_i(\mathbf{x}^*)+\sum_{j=1}^{p}\mu_j{}^*\nabla g_j(\mathbf{x}^*)=\mathbf{0},\ \ \mu_j{}^*\leq 0 \tag{56}$$

There is also a fourth case that is treated similarly.

In conclusion, we have 4 cases:

i) $\min f(\mathbf{x})$

subject to:

$h_i(\mathbf{x})=0,\ \ i=1,...,m$ and $g_j(\mathbf{x})\leq 0,\ \ j=1,...,p$.

ii) $\max f(\mathbf{x})$

subject to:

$h_i(\mathbf{x})=0,\ \ i=1,...,m$ and $g_j(\mathbf{x})\leq 0,\ \ j=1,...,p$.

iii) $\min f(\mathbf{x})$

subject to:

$h_i(\mathbf{x})=0,\ \ i=1,...,m$ and $g_j(\mathbf{x})\geq 0,\ \ j=1,...,p$.

iv) $\max f(\mathbf{x})$

subject to:

$h_i(\mathbf{x})=0,\ \ i=1,...,m$ and $g_j(\mathbf{x})\geq 0,\ \ j=1,...,p$.

If we write the KKT condition at point $\mathbf{x}^*$, for all 4 cases we get that:

$$\nabla L(\mathbf{x}^*, \lambda^*, \mu) = \tag{57}$$
$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j{}^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}$$

and $\mu_j{}^* g_j(\mathbf{x}^*) = 0$ for each $j = 1, \ldots, p$.
For each of the 4 cases, we have:
i) $\mu_j \geq 0, \;\; j = 1, ..., p$.
ii) $\mu_j \leq 0, \;\; j = 1, ..., p$.
iii) $\mu_j \leq 0, \;\; j = 1, ..., p$.
iv) $\mu_j \geq 0, \;\; j = 1, ..., p$.

Example:

Which are the conditions of KKT theorem for the problem below:

$$\min f(\mathbf{x}) \quad \text{subject to:} \quad \mathbf{x} \geq \mathbf{0}.$$

Solution

$$\nabla f(\mathbf{x}^*) \geq \mathbf{0}$$
$$\mathbf{x}^* \geq \mathbf{0} \text{ and}$$
$$\sum_i x_i \frac{\partial f(x^*)}{\partial x_i} = 0. \tag{58}$$

6.3 Second-order necessary conditions for existence of local minimum

If $\mathbf{x}^*$ is a regular point of the constraints:

$$h_i(\mathbf{x}^*) = 0, \;\; i = 1, ..., m$$
$$g_j(\mathbf{x}^*) \leq 0, \;\; j = 1, ..., p$$

and if $\mathbf{x}^*$ is local minimum of f subject to the constraints above, then $\exists\, \lambda^* \in \mathbb{R}^m, \;\; \mu^* \geq \mathbf{0} \;\; (\in \mathbb{R}^p_+)$ such that:

i) $\nabla L(\mathbf{x}^*, \lambda^*, \mu^*) = \mathbf{0}$

ii) $\mu_j{}^* g_j(\mathbf{x}^*) = 0, \; j = 1, \ldots, p$

iii) The Hessian matrix of the Lagrangian function,

$$\Lambda(\mathbf{x}^*, \lambda^*, \mu^*) = F(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* H_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j{}^* G_j(\mathbf{x}^*) \tag{59}$$

where $F(\mathbf{x}^*)$: Hessian matrix of $f(\mathbf{x})$ at $\mathbf{x}^*$,
$H_i(\mathbf{x}^*)$: Hessian matrix of $h_i(\mathbf{x})$ at $\mathbf{x}^*$,
$G_j(\mathbf{x}^*)$: Hessian matrix of $g_j(\mathbf{x})$ at $\mathbf{x}^*$
is positive semi-definite on the tangent subspace of the active constraints at $\mathbf{x}^*$.

6.4 Second Order Sufficient Conditions for existence of local minimum

Consider the problem:

$$\min f(\mathbf{x})$$

$$\text{s.t. } h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m,$$
$$g_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, p.$$

If $\exists \quad \boldsymbol{\lambda^*} \in R^m,\ \boldsymbol{\mu^*} \in R_+^p,\ \ (ie.\boldsymbol{\mu} \geq 0)$ such that:

$$\nabla L(\mathbf{x}^*, \boldsymbol{\lambda^*}, \boldsymbol{\mu}^*) = \tag{60}$$

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j{}^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}$$

and

$$\mu_j g_j(\mathbf{x}^*) = 0, \quad j = 1, \ldots, p \tag{61}$$

and

The Hessian matrix of the Lagrangian function,

$$\Lambda(\mathbf{x}^*, \boldsymbol{\lambda^*}, \boldsymbol{\mu^*}) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* H_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j{}^* G_j(\mathbf{x}^*) > 0 \tag{62}$$

is positive-definite on the subspace:

$$\mathcal{M} = \{\mathbf{y} : \nabla h_i(\mathbf{x}^*)^T \mathbf{y} = 0, \nabla g_j(\mathbf{x}^*)^T \mathbf{y} = 0 \text{ for } j \in \mathcal{J}(\mathbf{x}^*)\}, \tag{63}$$

with $\mathcal{J}(\mathbf{x}^*) = \{j : g_j(\mathbf{x}^*) = 0, \mu_j{}^* > 0\}$ the set of active constraints at point $\mathbf{x}^*$. Thus, $\mathcal{M}$ is the subspace that is tangent level to the surface of the active constraints
then $\mathbf{x}^*$ is local optimum of function f.

Remark: Note that in our problems, unless otherwise specified, we will always assume that $\mathcal{M} = \mathbb{R}^m$ and thus we will not worry about finding an explicit characterization of $\mathcal{M}$. However, we will need to show that matrix $\Lambda(\cdot)$ is positive-definite, i.e for all $\mathbf{y} \in \mathbb{R}^m$ it is $\mathbf{y}^T \Lambda \ \mathbf{y} > 0, \ \ \mathbf{y} \neq 0$.

6.5 Second-order Sufficient Conditions for Convex Functions

Consider the problem:

$$\text{min } f(\mathbf{x})$$

$$\text{subject to:}$$
$$\begin{array}{ll} h_i(\mathbf{x}) = 0, & i = 1, \ldots, m \\ g_j(\mathbf{x}) \leq 0, & j = 1, \ldots, p, \end{array}$$

with functions f, h_i, g_j convex, $i = 1, \ldots, m$ and $j = 1, \ldots, p$. The second-order sufficient conditions for existence of minimum in this case are as follows:

If $\exists \quad \mathbf{x}^*, \boldsymbol{\lambda^*} \in R^m, \ \boldsymbol{\mu^*} \in R^p_+. \ \boldsymbol{\mu^*} \geq 0$ such that:

$$\nabla L(\mathbf{x}^*, \boldsymbol{\lambda^*}, \boldsymbol{\mu^*}) = \tag{64}$$
$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i{}^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j{}^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}$$

and

$$\mu_j{}^* g_j(\mathbf{x}^*) = 0, \quad j = 1, \ldots, p \tag{65}$$

then $\mathbf{x}^*$ is global minimum of function f subject to the constraints.

7 Sensitivity analysis

Consider the problem:

$$\text{min } f(\mathbf{x})$$

$$\text{s.t. } \mathbf{h}(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

Note that we have collectively described all equality constraints and all inequality constraints with two vectors $\mathbf{h}(\cdot)$ and $\mathbf{g}(\cdot)$ respectively. Now, assume we increase the right-hand side of the constraints (resources) as follows:

$$\mathbf{h}(\mathbf{x}) = \mathbf{c}$$
$$\mathbf{g}(\mathbf{x}) \leq \mathbf{d}$$

Let $\mathbf{x}(\mathbf{0}, \mathbf{0})$ to be the optimal solution to the initial problem and let $\mathbf{x}(\mathbf{c}, \mathbf{d})$ be the solution of the problem formed after we increased the right-hand sides of constraints. Then by following a similar reasoning as the one in the sensitivity analysis for problems with equality constraints, we have:

$$\nabla_{\mathbf{c}} f(\mathbf{x}, (\mathbf{c}, \mathbf{d})) = -\boldsymbol{\lambda}$$
$$\nabla_{\mathbf{d}} f(\mathbf{x}, (\mathbf{c}, \mathbf{d})) = -\boldsymbol{\mu}$$

and for the i-th Lagrange multiplier we have:

$$\lambda_i = -\frac{\partial f(\mathbf{x}(\mathbf{c},\mathbf{d}))}{\partial c_i}\Bigg|_{(\mathbf{c},\mathbf{d})=(\mathbf{0},\mathbf{0})}$$

namely λ_i is the rate of change of the objective function with respect to a unit of change in the i-th equality constraint, i.e it is the derivative of the cost function with respect to the quantity c_i that the i-th equality constraint changes.
For the j-th KKT multiplier we have:

$$\mu_j = -\frac{\partial f(\mathbf{x}(\mathbf{c},\mathbf{d}))}{\partial d_j}\Bigg|_{(\mathbf{c},\mathbf{d})=(\mathbf{0},\mathbf{0})}$$

namely μ_j is the rate of change of the objective function with respect to a unit of change in the j-th inequality constraint, i.e it is the derivative of the cost function with respect to the quantity d_j that the j-th inequality constraint changes.

Thus, λ_i, μ_j can be interpreted as price per unit of the corresponding resource that is described by the i-th equality or the j-th inequality constraint.

7.1 Problem

Consider the problem

$$\text{min} f(x_1, x_2) = (x1 - 1)^2 + x_2 - 2$$

$$\text{subject to: } h(\mathbf{x}) = x_2 - x_1 - 1 = 0 \quad (\lambda)$$

$$g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0 \quad (\mu \geq 0)$$

We define one Lagrange multiplier λ for the equality constraint and one KKT multiplier μ for the inequality constraint. We define the Lagrangian function and we consider:

$$\nabla L(x_1, x_2, \lambda, \mu) = \mathbf{0} \Rightarrow \begin{cases} \frac{\partial L(\cdot)}{\partial x_1} = 0 \\ \frac{\partial L(\cdot)}{\partial x_2} = 0 \end{cases}$$

and also we have the condition:

$\mu(x_1 + x_2 - 2) = 0$
$\mu \geq 0$

We have the following two cases:

a) $\mu > 0 \Rightarrow x_1 + x_2 - 2 = 0$
Thus, using the 4 equations we compute $x_1^* = 1/2, x_2^* = 3/2, \lambda^* = -1, \mu^* = 0$ and we confirm whether $\mu > 0$ is satisfied. Since $\mu = 0$ we arrive at paradox so this is not the case, and we proceed to the next case.

b) $\mu = 0 \Rightarrow x_1 + x_2 - 2 < 0$
Thus, using the 3 equations we compute $x_1^* = 1/2, x_2^* = 3/2, \lambda^* = -1$ and we confirm whether (x_1^*, x_2^*) feasible. It turns out that this is the case, and thus (x_1^*, x_2^*) is the optimal solution to the problem with
$f(x_1^*, x_2^*) = \frac{3}{4}$.

Note that if both cases produced a feasible solution, then the optimal would be the one that minimizes the objective's value. Also, observe that we didn't use the second order condition because the primal function is convex and thus Hessian matrix is, without doubt, positive-definite.

8 The water-filling algorithm

As an example of an optimization problem with equality and inequality constraints, we will consider a problem that arises in various resource allocation instances. Consider a transmitter that has at its disposal N orthogonal channels for transmission. The transmitter has a total amount of power P, assume that $P = 1$ without loss of generality.

The transmission rate (capacity) C_i for a channel i is given by $C_i = \log_2(1 + \frac{P_i}{N_i})$, where P_i is transmission power assigned to channel i and N_i is the noise power (noise variance) of channel i, $i = 1, \ldots, N$.

The objective is to allocate (split) the available power across channels so as to maximize total achieved capacity in all channels. This problem arises in OFDM systems in which the transmitter transmits in parallel using N sub-carrier frequencies. Also, the capacity can be viewed as a special case of a utility function $U(\cdot)$. The utility function $U(x)$ measures the amount of satisfaction of a user or consumer if an amount x of good (power, bandwidth, etc) is allocated to it. Here, obviously the resource is the power and the capacity is the derived utility. The problem can be formulated as follows:

$$\max \sum_{i=1}^{N} \log_2 \left(1 + \frac{P_i}{N_i}\right)$$

$$\text{subject to: } \sum_{i=1}^{N} P_i = 1, \; P_i \geq 0, i = 1, \ldots, N. \qquad (66)$$

Note that in the most general case of utility function, we have the optimization problem of distributing a total amount of good W across users or consumers so as to maximize total utility

$$\max \sum_{i=1}^{N} U_i(x_i)$$

subject to

$$\sum_{i=1}^{N} x_i = W, \text{ and } x_i \geq 0, \qquad (67)$$

where $U_i(\cdot)$ is the utility function of user i.

Now, the capacity maximization problem we are dealing with can be written equivalently as

$$\min - \sum_{i=1}^{N} \log \left(1 + \frac{P_i}{N_i}\right)$$

subject to

$$\sum_{i=1}^{N} P_i = 1 \tag{68}$$

and

$$-P_i \leq 0, \;\; i = 1, \ldots, N \tag{69}$$

Define a Lagrange multiplier λ for the equality constraint and a KKT multiplier $\mu_i \geq 0$ for each inequality constraint $i = 1, \ldots, N$. Note also that in the above formulation we have omitted for simplicity the base 2 of the logarithm.

The initial problem is actually case 4 of the four cases we considered in the previous lecture, while the transformed one is of the form of case 1. Both are equivalent. The Lagrangian function is,

$$\begin{aligned} L(\mathbf{P}, \lambda, \boldsymbol{\mu}) = & \\ & -\sum_{i=1}^{N} \log\left(1 + \frac{P_i}{N_i}\right) + \lambda\left(\sum_{i=1}^{N} P_i - 1\right) + \sum_{i=1}^{N} \mu_i(-P_i) \\ & = -\sum_{i=1}^{N} \log\left(1 + \frac{P_i}{N_i}\right) + \lambda\left(\sum_{i=1}^{N} P_i - 1\right) - \sum_{i=1}^{N} \mu_i P_i \end{aligned} \tag{70}$$

We apply KKT conditions to solve the problem:

$$\nabla_{\mathbf{P}} L(\mathbf{P},\lambda,\boldsymbol{\mu}) = \mathbf{0} \quad \Rightarrow \quad \frac{\partial L(\mathbf{P},\lambda,\boldsymbol{\mu})}{\partial P_i} = 0 \tag{71}$$

$$\Rightarrow \quad -\frac{1}{1+\frac{P_i}{N_i}}\frac{1}{N_i} + \lambda - \mu_i = 0 \; \forall\, i$$

and

$$\mu_i P_i = 0, \;\; \forall\, i \tag{72}$$

and we also have: $P_i \geq 0$, $\sum_{i=1}^{N} P_i = 1$ and $\mu_i \geq 0$, $i = 1, \ldots, N$.

We solve equation (71) for μ_i:

$$\mu_i = \lambda - \frac{1}{P_i + N_i} \tag{73}$$

and substitute in $\mu_i P_i = 0$ to get:

$$\left(\lambda - \frac{1}{P_i + N_i}\right) P_i = 0 \tag{74}$$

We distinguish three cases:

i) $P_i > 0$ and $\left(\lambda - \frac{1}{P_i+N_i}\right) P_i = 0$. Then,

$$\lambda = \frac{1}{P_i + N_i} \Rightarrow P_i = \frac{1}{\lambda} - N_i \tag{75}$$

However, since we do not know λ, we do not know the sign of $\frac{1}{\lambda} - N_i$, and we can say that if $\frac{1}{\lambda} > N_i \Rightarrow P_i = \frac{1}{\lambda} - N_i$, since it has to be $P_i \geq 0$.

ii) $\left(\lambda - \frac{1}{P_i+N_i}\right) > 0$ and $P_i = 0$. Then,

$$P_i = 0 \text{ if } \frac{1}{\lambda} < N_i \tag{76}$$

iii) $\frac{1}{\lambda} = N_i$. Then,

$$\left(\frac{1}{N_i} - \frac{1}{P_i + N_i}\right) P_i = 0 \Rightarrow \tag{77}$$

$$\frac{P_i + N_i - N_i}{(P_i + N_i)N_i} P_i = 0 \Rightarrow P_i = 0 \tag{78}$$

Thus in conclusion we have:

$$P_i^* = \begin{cases} \frac{1}{\lambda^*} - N_i, & \text{if } \dfrac{1}{\lambda^*} > N_i \\ 0, & \text{if } \frac{1}{\lambda^*} \leq N_i \end{cases}$$

or equivalently,

$$P_i^* = \left(\frac{1}{\lambda^*} - N_i\right)^+ = \max\left(0, \frac{1}{\lambda^*} - N_i\right) \tag{79}$$

$$\text{with} \quad x^+ = \begin{cases} x & \text{, if } x > 0 \\ 0 & \text{, if } x \leq 0 \end{cases}$$

From the form of the solution above, we can see that the quantity $\frac{1}{\lambda^*}$ is common for all channels and resembles a kind of ”water-level”. Note that the better quality the channel is (the smaller the noise power), the more power is allocated to it. Also, the more noisy the channel, the less the power that is allocated to it. If the channel is ”too noisy”, i.e the noise power exceeds a certain power, then it is better from a capacity point of view not to allocate any power in the channel.

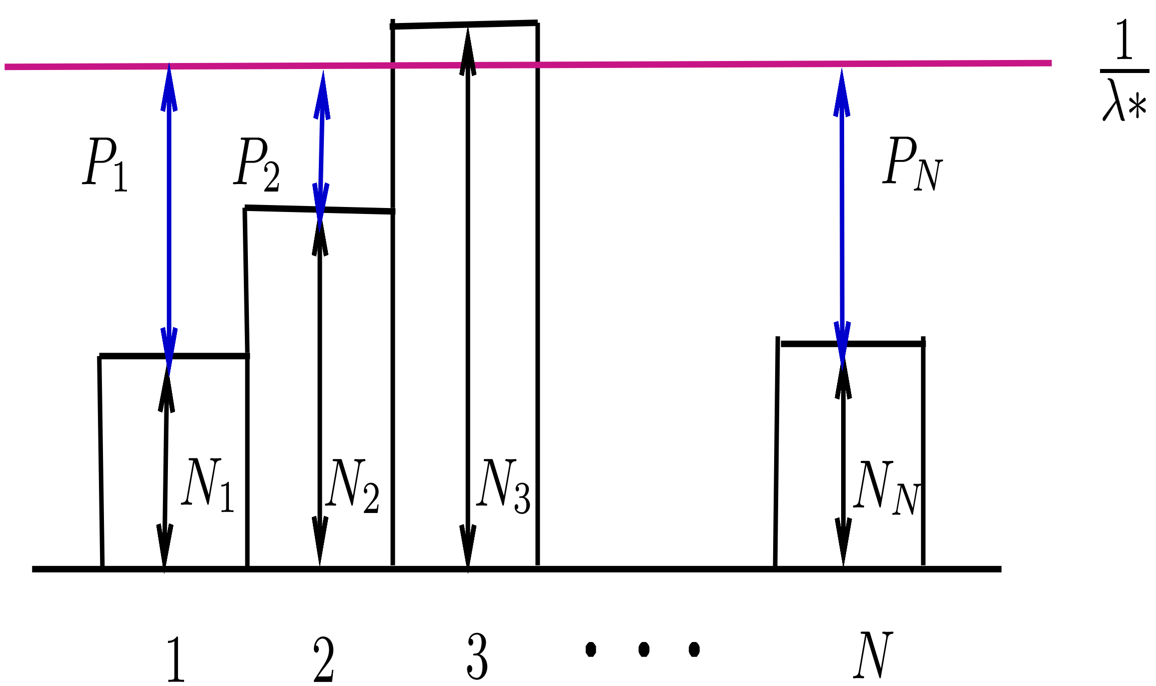

Figure 7: Water-filling algorithm with different amount of power P_i allocated to each channel i.

The result of water-filling can be seen in figure 7. In order to compute λ, one could argue that the constraint $\sum_{i=1}^{N} P_i^* = 1$ could be used, or $\sum_{i=1}^{N} \left(\frac{1}{\lambda^*} - N_i\right)^+ = 1$. However, it is not possible to solve the equation with regard to λ analytically, since we do not know in advance whether in different channels i the quantity $\frac{1}{\lambda^*} - N_i$ will be positive or negative.

Instead, we use a simple algorithm. Each column represents a channel and the height of each column reflects the noise. There are N channels that are differentiated due to different noise level. Also the quantity $\frac{1}{\lambda^*}$ as we said is common for all channels.

We divide available power $P = 1$ into small quantities ϵ. We start allocating power in small quantities ϵ to the channel i_1 with the best quality (the less noise power N_{i_1}) until we reach the level of a channel i_2 of the second best quality, i.e second smallest N. From that point, we assign power ϵ to each of channels i_1, i_2 until we reach channel i_3 with the third best quality (third smallest N). Then we allocate power ϵ to these three channels. We continue in that fashion until we exhaust the available power. The point where we exhaust the power defines the final water level $\frac{1}{\lambda^*}$. The procedure is shown in figure 8.

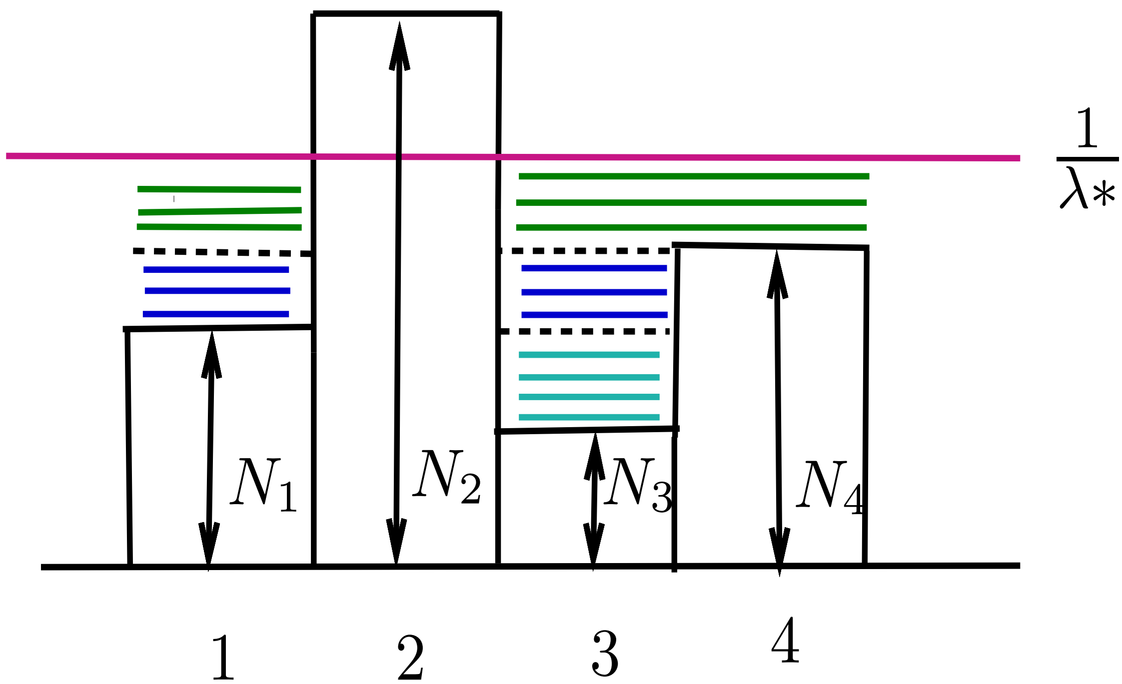

Figure 8: Successive power allocation in water-filling.

9 Lagrangian Duality in NLP

9.1 The Lagrangian

We consider an optimization problem (P) in the form:

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ s.t. & h_i(\mathbf{x}) = 0, & i = 1, \dots, p \\ & g_j(\mathbf{x}) \leq 0, & j = 1, \dots, m \end{array} \tag{80}$$

$$\Omega = \left\{ \begin{array}{ll} \mathbf{x}: & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \\ & g_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, m \end{array} \right\}$$

with variable $\mathbf{x} \in \Omega$.

In Lagrangian duality, we start by writing the Lagrangian:

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j g_j(\mathbf{x}) \quad (81)$$

We refer to λ_i as the *KKT multiplier* associated with the *i*th inequality constraint $f_i(\mathbf{x}) \leq 0$; similarly we refer to μ_i as the Lagrange multiplier associated with the *i*th equality constraint $h_i(\mathbf{x}) = 0$. Vectors λ and μ are called the *dual variable* vectors associated with the problem (P). We may call both kinds of variables Lagrangian variables.

9.2 The Lagrangian dual function

We define the *Lagrange dual function* l_D as the minimum value of the Lagrangian over x. That is, for $\lambda \in \mathbf{R}^m, \mu \in \mathbf{R}^p$,

$$\begin{aligned} l_D(\lambda, \mu) &= \min_{\mathbf{x} \in \Omega} L(\mathbf{x}, \lambda, \mu) \quad (82) \\ &= \min_{\mathbf{x} \in \Omega} \{ f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j g_j(\mathbf{x}) \} \end{aligned}$$

9.3 Lower bounds on optimal value

The Lagrangian dual function yields lower bounds on the optimal value $f(\mathbf{x}^*)$ of the problem (P): For any $\lambda \geq 0$ and any μ we have

$$l_D(\lambda, \mu) \leq f(\mathbf{x}^*). \tag{83}$$

Suppose $\mathbf{x}_0$ is a feasible point for the problem (1), i.e., $h_i(\mathbf{x}) = 0$, $g_j(\mathbf{x}) \leq 0$, and $\lambda \geq 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j g_j(\mathbf{x}_0) \leq 0 \tag{84}$$

since each term in the first sum is zero, and each term in the second sum is non-positive, and therefore

$$L(\mathbf{x}_0, \lambda, \mu) = f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j g_j(\mathbf{x}_0) \leq f(\mathbf{x}_0) \tag{85}$$

Note that in the second sum, we have that each term is non-positive, since $\mu_j \geq 0$ and $g_j(\mathbf{x}_0) \leq 0$. Thus,

$$l_D(\lambda, \mu) = \min_{\mathbf{x} \in \Omega} L(\mathbf{x}, \lambda, \mu) \leq L(\mathbf{x}_0, \lambda, \mu) \leq f(\mathbf{x}_0) \tag{86}$$

So $\forall\ \mathbf{x}_0$ which is feasible, $l_D(\lambda, \mu) \leq f(\mathbf{x}_0) \leq f(\mathbf{x}^*)$, and (83) holds.

9.4 Lagrangian dual problem

We saw from inequality (83), that we can get a lower bound from l_D on the optimal value of the objective function for each pair $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ with $\boldsymbol{\mu} \geq \mathbf{0}$. The next natural question is to find the *best* lower bound that can be obtained by the Lagrangian dual function. This brings us to the formulation of the Lagrangian dual problem (LD), which is given as:

$$\begin{array}{ll} \max & l_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ s.t. & \boldsymbol{\mu} \geq \mathbf{0}. \end{array} \tag{87}$$

There are two main reasons why we prefer to solve (LD) problem instead of the original one (P). The first reason is because it may be easier to solve (LD) since it has fewer constraints, and second and most important, because $l_D(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is always concave, independently of the original (P). This last claim can be shown by the following line of thoughts:

Consider functions of one variable for simplicity. First we need to show that if $f_1(x), f_2(x)$ are concave functions, then $f(x) = \min\{f_1(x), f_2(x)\}$ is also concave.
We know that a function is concave when:
$f(\vartheta x + (1-\vartheta)y) \geq \vartheta f(x) + (1-\vartheta) f(y)$.
In our case:

$$
\begin{aligned}
&f(\vartheta x + (1-\vartheta)y) = \\
&\qquad \min\{f_1(\vartheta x + (1-\vartheta)y), f_2(\vartheta x + (1-\vartheta)y)\} \geq \\
&\qquad \min\{\vartheta f_1(x) + (1-\vartheta)f_1(y), \vartheta f_2(x) + (1-\vartheta)f_2(y)\} \geq \\
&\qquad \vartheta \min\{f_1(x), f_2(x)\} \;+\; (1-\vartheta)\min\{f_1(y), f_2(y)\} = \\
&\qquad \vartheta f(x) + (1-\vartheta)f(y)
\end{aligned}
$$

The above result can be extended for more functions $f_1(x), f_2(x), \ldots, f_n(x)$ and is also valid for convex functions $f_1(\cdot), f_2(\cdot)$, if $\min$ is substituted by $\max$. Namely, we can show that if $f_1(\cdot), f_2(\cdot)$ are convex, then $\max\{f_1(\cdot), f_2(\cdot)\}$ is a convex function.

So from the (LD) problem we see that $l_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x}}\{$ linear functions of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}\}$. Since we know that all linear functions can be considered to be concave, the proof is completed.

9.5 Week Duality

The optimal value of the Lagrangian dual problem (LD), which we denote $d^* = l_D(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, is, by definition, the best lower bound on $p^* = f(\mathbf{x}^*)$, which is the optimal value of primal problem (P). In particular, we have the important inequality:

$$d^* \leq p^*. \tag{88}$$

This property is called *Weak Duality Lemma* and can be shown as follows:

$$f(\mathbf{x}) \geq p^* \tag{89}$$

where

$$\begin{aligned} p^* &= \min_{\mathbf{x}\in\Omega} f(\mathbf{x}) \\ &\geq \min_{\mathbf{x}\in\Omega}\{f(\mathbf{x}) + \sum_{i=1}^{m}\lambda_i h_i(\mathbf{x}) + \sum_{j=1}^{p}\mu_j g_j(\mathbf{x})\} = l_D(\lambda, \mu) \end{aligned}$$

$$\Rightarrow f(\mathbf{x}) \geq p^* \geq l_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) \ \forall \boldsymbol{\lambda}, \boldsymbol{\mu} \tag{90}$$

$$\Rightarrow f(\mathbf{x}) \geq p^* \geq l_D(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = d^* \tag{91}$$

The weak duality inequality (88) holds even if d^* and p^* are infinite. If the primal problem (P) is unbounded from below, so that $p^* = -\infty$, we must have $d^* = -\infty$, *i.e.*, the Lagrange dual problem is infeasible. Conversely, if the dual problem (LD) is unbounded from above, so that $d^* = \infty$, we must have $p^* = \infty$, *i.e.*, the primal problem is infeasible.

We refer to the difference $p^* - d^*$ as the *optimal duality gap* of the original problem, since it gives the gap between the optimal value of the primal problem and the best *i.e.* greatest lower bound on it that can be obtained from the Lagrangian dual function. The optimal duality gap is always non-negative.

9.6 Strong Duality

If the equality

$$d^* = p^* \tag{92}$$

holds, *i.e.*, the optimal duality gap is zero, then we say that *strong duality* holds. This means that the best bound that can be obtained from the Lagrange dual function is tight.

Strong duality does not, in general, hold. But if the primal problem (P) is convex, *i.e.*, of the form

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ s.t. & g_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, m \\ & A\mathbf{x} = \mathbf{b} \end{array} \tag{93}$$

with functions $f(\cdot), g_1(\cdot), \ldots, g_m(\cdot)$ convex, we usually (but not always) have strong duality. There are many results that establish conditions on the problem, beyond convexity, under which strong duality holds. These conditions are called constraint qualifications.

One such simple constraint qualification is *Slater's condition*. The condition says:
if there exists $\mathbf{x} \in \Omega$ such that

$$g_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, m, \qquad A\mathbf{x} = \mathbf{b} \tag{94}$$

then we have strong duality.

Note that in *Linear Programming* strong duality always holds.

9.7 Solving primal problem (P) using Lagrangian dual (LD)

We transform our primal problem (P) in its Lagrangian dual form

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\mu} \geq \mathbf{0}} l_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) \tag{95}$$

In order to find $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$ we start from an arbitrary $\boldsymbol{\lambda}_0$ and $\boldsymbol{\mu}_0$ and use the gradient ascent method:

$$\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \alpha \nabla_{\boldsymbol{\lambda}} l_D(\boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) \tag{96}$$

and

$$\boldsymbol{\mu}^{(t+1)} = \boldsymbol{\mu}^{(t)} + \alpha \nabla_{\boldsymbol{\mu}} l_D(\boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)). \tag{97}$$

If $\boldsymbol{\mu} < \mathbf{0}$ somewhere, then we substitute with $\mathbf{0}$.
In case $l_D(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is not differentiable, we use the so called super-gradient method. Then, denote a vector $\mathbf{w}$ as the super-gradient of $l_D(\cdot)$ the iteration becomes:

$$\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \alpha \cdot \mathbf{w}(t) \tag{98}$$

Also, in the same spirit, we can write the iteration for $\boldsymbol{\mu}$ using the corresponding supergradient.Note that this supergradient will be different than $\mathbf{w}$.

In general, vector $\mathbf{w}$ is called super-gradient of function f at $\mathbf{x}_0$ if and only if:

$$f(\mathbf{x}) - f(\mathbf{x}_0) \leq \mathbf{w}^T(\mathbf{x} - \mathbf{x}_0). \tag{99}$$

In our case $f(\mathbf{x}) \equiv l_D(\boldsymbol{\lambda}, \boldsymbol{\mu})$, so the above equation becomes:

$$l_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) - l_D(\boldsymbol{\lambda}(t), \boldsymbol{\mu}) \leq \mathbf{w}^T(\boldsymbol{\lambda} - \boldsymbol{\lambda}(t)) \tag{100}$$

Note: In the case that l_D is not differentiable, we use *super-gradient* only if the search direction is ascent. If search direction is descent we use the *sub-gradient*.
These topics are considered advanced and we will not elaborate in them more.

10 Saddle-point interpretation

In this section we give several interpretations of Lagrangian duality.

10.1 Max-min characterization of weak and strong duality

To simplify the discussion we assume there are no equality constraints. The results are easily extended to cover them. First note that

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\mu}) = \max_{\boldsymbol{\mu} \geq \mathbf{0}} \left(f(\mathbf{x}) + \sum_{j=1}^{p} \mu_j g_j(\mathbf{x}) \right) \tag{101}$$

We can express the optimal value of the primal problem as

$$p^* = \min_{\mathbf{x}} \max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\mu}). \tag{102}$$

By definition of the dual function, we also have

$$d^* = \max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}). \tag{103}$$

Thus, weak duality can be expressed as the inequality:

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) \leq \min_{\mathbf{x}} \max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\mu}) \tag{104}$$

and strong duality as the equality:

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) = \min_{\mathbf{x}} \max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\mu}) \tag{105}$$

Strong duality means that the order of the minimization over $\mathbf{x}$ and the maximization over $\boldsymbol{\mu} \geq \mathbf{0}$ can be switched without affecting the result.

In fact, the inequality (104) does not depend on any properties of function $L(\cdot)$, and therefore we have:

$$\max_{\mathbf{z} \in \mathcal{Z}} \min_{\mathbf{w} \in \mathcal{W}} f(\mathbf{w}, \mathbf{z}) \leq \min_{\mathbf{w} \in \mathcal{W}} \max_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{w}, \mathbf{z}) \tag{106}$$

for any $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ (and any $\mathcal{W} \subseteq \mathbf{R}^n$ and $\mathcal{Z} \subseteq \mathbf{R}^m$). This general inequality is called the *max-min inequality*. When equality holds, *i.e.*,

$$\max_{\mathbf{z} \in \mathcal{Z}} \min_{\mathbf{w} \in \mathcal{W}} f(\mathbf{w}, \mathbf{z}) = \min_{\mathbf{w} \in \mathcal{W}} \max_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{w}, \mathbf{z}) \tag{107}$$

we say that f (and $\mathcal{W}$ and $\mathcal{Z}$) satisfy the *strong max-min property* or the *saddle-point property*.

We refer to a pair $\mathbf{w}_0 \in \mathcal{W}, \mathbf{z}_0 \in \mathcal{Z}$ as a *saddle-point* of function $f(\cdot)$ (and $\mathcal{W}$ and $\mathcal{Z}$) if

$$f(\mathbf{w}_0, \mathbf{z}) \leq f(\mathbf{w}_0, \mathbf{z}_0) \leq f(\mathbf{w}, \mathbf{z}_0) \tag{108}$$

for all $\mathbf{w} \in \mathcal{W}$ and $\mathbf{z} \in \mathcal{Z}$.

10.2 Game theory interpretation of saddle point

We can interpret *max-min* inequality (104) the *max-min* equality (105) and the *saddle-point* property in terms of a *zero-sum game* with two players, where each players has a continuous set of strategies. If the first player (P1) chooses $\mathbf{w} \in \mathcal{W}$, and the second player (P2) selects $\mathbf{z} \in \mathcal{Z}$, then P1 pays an amount $f(\mathbf{w}, \mathbf{z})$ to P2. P1 therefore wants to minimize f, while P2 wants to maximize f. In this case there are certain connections between the saddle-point of $f(\cdot)$ and the so-called *Nash equilibrium point* of the game. If we are at the Nash equilibrium, no player can do better by deviating his strategy from that defined by the equilibrium.

11 Decomposition theory

11.1 Example : Network pricing

Consider a (wired) network, represented as a directed graph $G = (S, A)$ where S is the set of network nodes that belong at the network ($|S|$ is the number of nodes) and A is the set of links of the network ($|A|$ is the number of links). The capacity (in bits/sec) of link $l \in A$ is defined by c_l and shows the maximum number of bits per second that can be carried over the link. We define the variable x_s (in bits/sec) as the information generation rate at every node (source) of the network as well as the utility function $U_s(x_s)$, which varies for every node, and is concave function of x_s. Note that every node in the network can potentially be the information source (and thus belong in set S). The utility function $U_s(x_s)$ shows the amount of satisfaction derived by node s if it is allowed to transmit with rate x_s. The utility function is a concave function of x_s for each node and each node may have a different utility function.

We have seen in previous lectures the physical meaning of a function being concave. It means that the rate of satisfaction with regard to change in x_s is a decreasing function of x_s. That is, the user (node) is satisfied with a higher rate for small values of x_s and this satisfaction rate decreases as we assign more resources x_s to it. The capacity function $c(P) \sim \log(P)$ as a function of the allocated amount of power p is a special case of utility function.

Each node could be considered either as:

Source : the more information rate it sends to other nodes of the network, the more it is satisfied, or

Destination : the more information rate it receives, the more it is satisfied.

Suppose now a network consisting of many nodes. Our goal is to maximize the total utility of all nodes in the network,

$$\max_{\mathbf{x}} \sum_{s=1}^{S} U_s(x_s)$$

by appropriately controlling the information generation rate x_s. The vector of variables $\mathbf{x}$ is $\mathbf{x} = (x_1, x_2, \ldots, x_{|s|})$.

Suppose source $s \in S$ uses the set of links $L(s)$ in order to transfer the information it produces. Define for each link $l \in A$ as $S(l)$ to be the set of sources that use link l, i.e that transfer their information through that link. This makes obvious the first, and essentially the only constraint of our problem which is the fact that capacity of each link should not be exceeded,

$$\sum_{s:l \in L(s)} x_s \leq c_l, \ l = 1, \ldots, |A|$$

Every source sends its information through specific paths. The above problem is an optimization problem with inequality constraints, one for each link. In order to find the solution, we use the Lagrangian function. So we define coefficient $\lambda_l \geq 0$ as the KKT multipliers that correspond to link l, $l = 1, \ldots, |A|$ and we have:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{s=1}^{|S|} U_s(x_s) + \sum_{l=1}^{|A|} \lambda_l (c_l - \sum_{s:l \in L(s)} x_s)$$

$$= \sum_{s=1}^{|S|} U_s(x_s) + \sum_{l=1}^{|A|} \lambda_l c_l - \sum_{l=1}^{|A|} \lambda_l \sum_{s:l \in L(s)} x_s$$

$$= \sum_{s=1}^{|S|} U_s(x_s) - \sum_{s=1}^{|S|} \sum_{l=1}^{|A|} x_s \lambda_l + \sum_{l=1}^{|A|} \lambda_l c_l$$

$$= \sum_{s=1}^{|S|} \left[U_s(x_s) - x_s \sum_{l \in L(s)} \lambda_l \right] + \sum_{l=1}^{|A|} \lambda_l c_l$$

Suppose that the values of λ_l are known. It is possible for the global problem to be solved by each source individually, so that each source s finds the optimal rate as

$$x_s^* = \arg\max_{x_s} \left[U_s(x_s) - x_s \sum_{l \in L(s)} \lambda_l \right] \tag{109}$$

Hence, the initial global objective is decomposed in that way into separate optimization problems, one for each source. If each source s finds the optimal rate x_s^*, then we have collectively the optimal solution $\mathbf{x}^*$ for the problem.

Now define the Lagrangian dual problem,

$$l_D(\boldsymbol{\lambda}) = \max_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{s=1}^{|S|} \max_{x_s} \left[U_s(x_s) - x_s \sum_{l \in L(s)} \lambda_s \right]$$

The value λ_l is the price paid by the source s that uses link l for each unit of flow that it sends through link l. As we see, the initial maximization decomposes into $|S|$ apart maximization problems, one for each source. Each source s solves a separate problem of maximizing the net benefit, i.e the derived utility minus the total cost of using the links in set $L(s)$,

$$\max_{x_s} \left[U_s(x_s) - x_s \sum_{l \in L(s)} \lambda_l \right]$$

Each source s can compute the optimal solution, x_s^* as the root of the equation

$$\frac{dU(x_s)}{dx_s} - \sum_{l \in L(s)} \lambda_l = 0 \Rightarrow x_s^*(\boldsymbol{\lambda})$$

for a given price vector $\boldsymbol{\lambda}$.

The Lagrangian dual problem is:

$$\min_{\boldsymbol{\lambda}\geq\mathbf{0}} l_D(\boldsymbol{\lambda}) = \min_{\boldsymbol{\lambda}\geq\mathbf{0}} \left[\sum_{s=1}^{|S|} U_s(x_s^*) - x_s^*(\boldsymbol{\lambda}) \sum_{l\in L(s)} \lambda_l \right] \tag{110}$$

In the problem setup, there is a central agent (e.g the network price controller that is ran by the network operator) that finds the price of using each link depending on its use and popularity. Thus, it computes different prices λ_l for each link l. Then, it adapts the price using the following intuitive rule : whenever a link is over-used its price has to be increased so as to discourage users from using it and thus reduce the link load. On the other hand, when a link is under-utilized, the price has to be reduced so as to make it more attractive to users to transfer their information through this link. Then, given the certain computed link price vector $\boldsymbol{\lambda}$, each source solves the separate maximization problem and finds the amount of traffic x_s^* to send through each link so as to maximize the net utility. The values of x_s are then sent to the central unit, which solves the Lagrangian dual problem in order to recalculate the prices.

Since the minimization problem of the Lagrangian dual cannot be solved analytically, the central entity can perform one iteration of the gradient descent algorithm to update the link prices. Thus, the price for link l is updated as follows:

$$\lambda_l(t+1) = \lambda_l(t) - a\frac{\partial l_D(\boldsymbol{\lambda})}{\lambda_l} \Rightarrow$$

$$\lambda_l(t+1) = \max\left\{\lambda_l(t) - a(c_l - \sum_{s\in S(l)} x_s^* \boldsymbol{\lambda}(t)), 0\right\}$$

where a is the step size for the gradient descent algorithm and we have taken care so that λ_l does not take negative values. Note that the equation above arose since

$$\frac{\partial l_D}{\partial \lambda_l} = c_l - \sum_{s:l\in L(s)} x_s^*(\boldsymbol{\lambda}(t)). \tag{111}$$

It is possible that this sum could overcome the value of c_l and then the central unit must increase the value of price $\lambda_l(t+1)$ at the next iteration. In contrast, when a link is not used very much, the value of $\lambda_l(t+1)$ decreases in order to make that link attractive and used by more sources.

11.2 Primal-Dual Algorithm

The Algorithm that takes place is as follows:

1. Start with initial prices $\lambda_l(0)$, for $l = 1, \ldots, |A|$.

2. Each source s solves, independently from the other sources,

the separate maximization problem

$$\max_{x_s} \left[U_s(x_s) - x_s \sum_{l \in L(s)} \lambda_l \right]$$

and finds the optimal rate $x_s^*(\lambda)$. Each source sends the optimal values $x_s^*(\lambda)$ to the central coordinating agent.

3. The central agent updates the price for using each link as follows:

$$\lambda_l(t+1) = \lambda_l(t) - a \frac{\partial l_D(\boldsymbol{\lambda})}{\lambda_l} \Rightarrow$$

$$\lambda_l(t+1) = \max\left\{ \lambda_l(t) - a(c_l - \sum_{s \in S(l)} x_s^* \boldsymbol{\lambda}(t)), 0 \right\}$$

and broadcasts the new link prices to all sources.

4. $t \leftarrow t+1$. Go to 1. Continue until convergence.

The algorithm above can be shown to converge to the optimal rate vector $\mathbf{x}^* = (x_1^*, \ldots, x_{|S|}^*)$, such that the total utility is maximized.

Note: We could view the price update mechanism as a form of congestion control for the reasons explained above.

12 Extra Notes

We have the primal problem:

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) \\ &s.t. \qquad g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \text{ and } \mathbf{x} \in \mathcal{X} \end{aligned} \tag{112}$$

Let

$$f^* = \inf_{\substack{\mathbf{x} \in X \\ g_i(\mathbf{x}) \leq 0 \\ i=1,\ldots,m}} f(\mathbf{x}) \tag{113}$$

Now define the Langrancian primal function:

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) \; + \; \sum_{i=1}^{m} \boldsymbol{\mu}_i g_i(\mathbf{x}) \tag{114}$$

Then the Langrangian dual function is:

$$l_D(\boldsymbol{\mu}) = \min_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\mu}) \tag{115}$$

The formulation of the Langrangian dual problem is:

$$\begin{aligned} &\text{maximize} \quad l_D(\boldsymbol{\mu}) \\ &s.t. \qquad \boldsymbol{\mu} \geq 0 \end{aligned} \tag{116}$$

Let

$$q^* = \sup_{\boldsymbol{\mu} \geq 0} l_D(\boldsymbol{\mu}) \tag{117}$$

So, for all $\boldsymbol{\mu} \geq 0,\ \mathbf{x} \in \mathcal{X},\ g(\mathbf{x}) \leq 0$, we have that:

$$l_D(\boldsymbol{\mu}) = \min_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\mu}) \leq f(\mathbf{x}) + \sum_{i=1}^{m} \boldsymbol{\mu}_i g_i(\mathbf{x}) \leq f(\mathbf{x})$$

Thus, we have:

$$q^* = \sup_{\boldsymbol{\mu} \geq 0} l_D(\boldsymbol{\mu}) \leq \inf_{\substack{\mathbf{x} \in X \\ g(\mathbf{x}) \leq 0}} f(\mathbf{x}) = f^* \qquad (118)$$

Example of dual problem

Consider the primal problem:

$$\begin{array}{ll} \text{min} & f(\mathbf{x}) = \frac{1}{2}\left({x_1}^2 + {x_2}^2\right) \\ s.t. & g(\mathbf{x}) = x_1 - 1 \leq 0, \qquad \mathbf{x} \in \mathcal{X} = \mathbb{R}^2 \end{array} \qquad (119)$$

The Langrangian function is:

$$L(\mathbf{x}, \boldsymbol{\mu}) = \frac{1}{2}\left({x_1}^2 + {x_2}^2\right) + \boldsymbol{\mu}\left(x_1 - 1\right), \quad \boldsymbol{\mu} \geq 0 \qquad (120)$$

We equate the partion derivatives $\frac{\partial L}{\partial x_i} = 0, \quad i = 1, 2$ and we get the equations:

$$\frac{\partial L}{\partial x_1} = 0 \quad \Rightarrow \quad x_1 + \boldsymbol{\mu} = 0 \qquad (121)$$

$$\frac{\partial L}{\partial x_2} = 0 \quad \Rightarrow \quad x_2 = 0 \qquad (122)$$

and we also have the constraint:

$$\boldsymbol{\mu}\left(x_1 - 1\right) = 0 \tag{123}$$

Using the above contraint, we observe that if we consider $\boldsymbol{\mu} > 0$ then $x_1 - 1 = 0 \Rightarrow x_1 = 1$ and thus $\boldsymbol{\mu} = -1$ which contradicts the initial assumption that $\boldsymbol{\mu} \geq 0$.
So, we have that $\boldsymbol{\mu} = 0$ and thus ${x_1}^* = {x_2}^* = 0$.
The Langrangian dual function is:

$$\begin{aligned} l_D(\boldsymbol{\mu}) &= \min_{\mathbf{x} \in \mathcal{R}^2} L(\mathbf{x}, \boldsymbol{\mu}) \\ &= \frac{1}{2}\left({x_1}^2 + {x_2}^2\right) + \boldsymbol{\mu}\left(x_1 - 1\right) \end{aligned} \tag{124}$$

Equating the partion derivatives $\frac{\partial L}{\partial x_i} = 0, \quad i = 1, 2$ we get that $x_1 = -\boldsymbol{\mu}$ and $x_2 = 0$. Then replacing these into the above equation we have:

$$l_D(\boldsymbol{\mu}) = -\frac{1}{2}\boldsymbol{\mu}^2 - \boldsymbol{\mu} \tag{125}$$

Now we define the dual problem:

$$\begin{array}{ll} \text{max} & l_D(\boldsymbol{\mu}) \\ s.t. & \boldsymbol{\mu} \geq 0 \end{array} \tag{126}$$

which gives us that $\boldsymbol{\mu}^* = 0$.
So, we can verify that there is *no duality gap* (remember that

the optimal duality gap is zero).

Examples

$$\begin{array}{ll} \text{min} & f(\mathbf{x}) = x_1 - x_2 \\ s.t. & g(\mathbf{x}) = x_1 + x_2 - 1 \leq 0 \\ & \mathbf{x} \in \mathcal{X} = \{(x_1, x_2) : x_1 \geq 0,\ \ x_2 \geq 0\} \end{array} \tag{127}$$

$$\begin{array}{ll} \text{min} & f(\mathbf{x}) = |x_1| + x_2 \\ s.t. & g(\mathbf{x}) = x_1 \leq 0 \\ & \mathbf{x} \in \mathcal{X} = \{(x_1, x_2 : x_2 \geq 0)\} \end{array} \tag{128}$$

Distributed implementation of waterfilling solution

$$\begin{array}{ll} \text{max} & \sum_{i=1}^{M} \log\left(1 + \frac{P_i}{N_i}\right) \\ s.t. & \sum_{i=1}^{M} P_i \leq P \end{array} \tag{129}$$

Sandle Point: $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is an optimal solution multiplier pair iff $\mathbf{x}^* \in \mathcal{X}, \boldsymbol{\mu}^* \geq 0$ and $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is saddle point of the Langrangian in the sense that

$$L(\mathbf{x}^*, \boldsymbol{\mu}) \leq L(\mathbf{x}^*, \boldsymbol{\mu}^*) \leq L(\mathbf{x}, \boldsymbol{\mu}^*) \tag{130}$$

$$f^* = l_D(\boldsymbol{\mu}^*) = \max_{\mu} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) \tag{131}$$

and

$$q^* = \min_{\mathbf{x}} \max_{\mu \geq 0} L(\mathbf{x}, \boldsymbol{\mu}) \tag{132}$$

If $q^* = f^*$ we have:

$$L(P, \boldsymbol{\mu}) = -\sum_{i=1}^{M} \log\left(1 + \frac{P_i}{N_i}\right) + \boldsymbol{\mu}\left(\sum_{i=1}^{M} P_i - P\right)$$
$$= \sum_{i=1}^{M}\left[-\log\left(1 + \frac{P_i}{N_i}\right) + \boldsymbol{\mu} P_i\right] - \boldsymbol{\mu} P$$

Given a Lagrange multiplier $\boldsymbol{\mu}_0$ each user separately tries to minimize the Lagrangian.Due to the separable character of the problem we have:

$$\frac{\partial\left[-\log\left(1 + \frac{P_i}{N_i}\right) + \boldsymbol{\mu} P_i\right]}{\partial P_i} = 0 \quad (133)$$
$$\Rightarrow \quad -\frac{1}{P_i + N_i} + \boldsymbol{\mu} = 0$$
$$\Rightarrow \quad P_i = \cdots$$

Given all the P_i's the BF finds the Langrange multipliers to maximize the L using the iteration:

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} + \left(\sum_{i=1}^{M} P_i^{(k)} - P\right) \quad (134)$$