Kuhn-Tucker conditions, Gambling and Investing

by Henk Tijms, 12 November 2021[1]

Abstract The goal of this tutorial is to give a first introduction to the Kelly betting system and to the Kuhn-Tucker conditions in nonlinear programming. The Kelly system is often used in the fields of gambling and investing, which fields have many similar traits. The formula will be derived both for the case of a single betting object and for the case of multiple betting objects, where the derivation for the latter case is based on the Kuhn-Tucker conditions from nonlinear programming. An elementary treatment of these famous conditions will be given in the Appendix of the article. The Kelly formula will be illustrated with betting examples from investment, soccer and horse races.

1. Introduction

In his book *A Mathematician Plays the Stock Market*, John Allen Paulos describes a scenario that occurred during the wild times when dotcom companies were going public on a daily basis. A certain investor is offered the following opportunity: Every Monday for a period of 52 weeks the investor may invest funds in the stock of one dotcom company. On the ensuing Friday, the investor sells. The following Monday, he purchases new stock in another dotcom company. Each week, the value of the stock purchased has a probability of $\frac{1}{2}$ of increasing by 80%, and a probability of $\frac{1}{2}$ of decreasing by 60%, independently of what happened in previous weeks. This means that on average, the increase in value of the purchased stock is equal to

$$0.8 \times \frac{1}{2} - 0.6 \times \frac{1}{2} = 0.1,$$

giving an average return of 10% per week. The investor, who has a starting bankroll of ten thousand dollars to invest over a period of the coming 52 weeks, doesn't hesitate for a moment; he decides to invest the full amount, every week, in the stock of a dotcom company. After 52 weeks, it appears that our investor only has 2 dollars left of his initial ten-thousand-dollar bankroll. He is, quite literally, at a loss to figure it all out. But in fact, this investment result is not very surprising when you consider how dangerous it is to rely on averages in situations involving uncertainty. A person can drown, after all, in a lake that has an average depth of 25 cm. For situations involving uncertainty factors, you should never work with averages, but rather with probabilities! It is easily explained that the probability of nearly depleting the bankroll is large if the investor invests his whole bankroll in each transaction. The most likely path to develop over the course of 52

[1]Henk Tijms is emeritus professor of operations research at the Vrije University in Amsterdam, email: h.c.tijms@xs4all.nl.

weeks is one in which the stock increases in value 50% of the time, and decreases in value 50% of the time. This path results in a bankroll of $1.8^{26} \times 0.4^{26} \times 10\,000 = 1.95$ dollars after 52 weeks. Running one hundred thousand simulations of the investments over 52 weeks renders a probability of about 50% that the investor's final bankroll will not exceed 1.95 dollars, and a small probability of 5.8% that the investor's final bankroll will be greater than his starting bankroll of ten thousand dollars.

Misled by seemingly favorable averages, our foolhardy investor stakes the full amount of his bankroll every week. Apparently, he is unacquainted with the *Kelly strategy.* According to this strategy, rather than investing the full amount of his current bankroll for every transaction, he would do better to invest a same fixed fraction of his current bankroll each time and to keep in reserve a fixed fraction of his current capital.

The Kelly model will be discussed in the next sections.[2] First, the case of a single betting object is analyzed and then that of multiple betting objects, as in a horse race with several horses. The appendix gives an introduction to the Kuhn-Tucker conditions that will be needed to analyze the case of multiple betting objects.

2. Kelly model with a single betting object

Consider the situation in which you can repeatedly make bets in a particular game with a single betting object. The game is assumed to be favorable to you, where favorable means that the expected value of the net payoff of the game is positive. For every dollar staked on a repetition of the game, you receive w_1 dollars back with probability p and w_2 dollars with probability $1-p$, where $0 < p < 1$, $w_1 > 1$ and $0 \leq w_2 < 1$. The outcomes of the successive games are assumed to be independent of each other. The key assumption for the Kelly betting model is

Assumption: *The parameters* p, w_1 *and* w_2 *satisfy* $pw_1 + (1-p)w_2 > 1$ *and* $pw_1 + (1-p)w_2 - 1 < (w_1 - 1)(1 - w_2)$.

The first condition says that the game is favorable to you in terms of one-step expected value. It is noted that the first condition implies the second condition if $w_2 = 0$. You start with a certain bankroll, and it is assumed

[2]This model is named after the physicist John Kelly Jr. Working at Bell Labs, he published in 1956 a paper titled *A New Interpretation of Information Rate* in the Bell System Technical Journal. Virtually no one took much note of the article when it first appeared. Nowadays it is widely used in gambling and investing. In the paper Kelly posited a scenario in which a horse-race better has an edge: a 'private wire' of somewhat reliable, but not perfect tips from inside information. How should he bet? Wager too little, and the advantage is squandered. Too much, and ruin beckons. The Kelly bet size is found by maximizing the expected value of the logarithm of wealth, which is equivalent to maximizing the expected geometric growth rate. In the context of the St. Petersburg paradox, it was already suggested by Daniel Bernoulli in 1738 that a gambler should not maximize expected return but rather logarithmic utility.

that you may stake any amount up to the size of your current bankroll each time. If you want to maximize the growth rate of your bankroll over the long run, the *Kelly formula* advises you to stake the following fixed fraction α of your current bankroll each time:

$$\alpha = \frac{pw_1 + (1-p)w_2 - 1}{(w_1 - 1)(1 - w_2)}. \tag{1}$$

This formula will be derived in the next section. Note that, by the assumption made, $0 < \alpha < 1$. In the special case of $w_2 = 0$, the Kelly formula (1) reduces to

$$\alpha = \frac{pw_1 - 1}{w_1 - 1}, \tag{2}$$

which can be interpreted as the ratio of the expected net gain per staked dollar and the payoff odds.

In the specific case of the investor with $p = 0.5$, $w_1 = 1.8$ and $w_2 = 0.4$, the Kelly strategy requires him to invest a fraction $\frac{5}{24}$ of his current bankroll for each transaction. In practical terms, this renders a practically zero probability of his ending with 1.95 dollars or less after 52 weeks. Simulation reveals that applying the Kelly strategy would give the investor about a 70% probability of ending with more than ten thousand dollars after 52 weeks, and about a 44% probability of ending with more than twenty thousand dollars.

The Kelly strategy was first used in casinos by mathematician Edward Thorp, in order to try out his winning blackjack system. Later, Thorp and a host of famous investors including Warren Buffett, successfully applied a form of the Kelly strategy to guide their stock market decisions.

3. **Derivation of the Kelly formula**

The strategy is to bet a fixed fraction α of your current bankroll each time, where $0 < \alpha < 1$. Here it is supposed that winnings are reinvested and that your bankroll is infinitely divisible. Letting V_0 be your starting bankroll, define the random variable V_m as

$$V_m = \text{the size of your bankroll after } m \text{ bets.}$$

For the mth bet, let the random variable W_m be equal to w_1 with probability p and be equal to w_2 with probability $1-p$. Noting that $V_m = (1-\alpha)V_{m-1} + \alpha V_{m-1}W_m$, it follows by induction that

$$V_m = \big(1 - \alpha + \alpha W_1\big) \times \cdots \times \big(1 - \alpha + \alpha W_m\big) V_0 \quad \text{for } m = 1, 2, \ldots.$$

In mathematics, a growth process is most often described with the help of an exponential function. This is the motivation to define the exponential growth factor G_m via the relationship

$$V_m = V_0 e^{mG_m},$$

where $e = 2.71828\ldots$ is the base of the natural logarithm. If you take the logarithm of both sides of this equation, you see that the definition of G_m is equivalent to

$$G_m = \frac{1}{m} \ln\left(\frac{V_m}{V_0}\right).$$

Using the product formula for V_m and the fact that $\ln(ab) = \ln(a) + \ln(b)$, you find

$$G_m = \frac{1}{m}\Big[\ln\big(1-\alpha+\alpha W_1\big) + \cdots + \ln\big(1-\alpha+\alpha W_m\big)\Big].$$

Next, we apply the law of large numbers, being one of the pillars of probability theory. Since the random variables $X_i = \ln(1-\alpha+\alpha W_i)$ form a sequence of independent random variables having a common probability distribution, the law of large numbers gives

$$\lim_{m\to\infty} G_m = E\big[\ln(1-\alpha+\alpha W)\big] \quad \text{with probability 1,}$$

where the random variable W is equal to w_1 with probability p and is equal to w_2 with probability $1-p$. Thus the long-run growth rate of your bankroll is equal to

$$\lim_{m\to\infty} G_m = p\ln(1-\alpha+\alpha w_1) + (1-p)\ln(1-\alpha+\alpha w_2) \quad \text{with probability 1.}$$

Putting the derivative of $g(\alpha) = p\ln(1-\alpha+\alpha w_1) + (1-p)\ln(1-\alpha+\alpha w_2)$ equal to 0, you get

$$\frac{p(w_1-1)}{1-\alpha+\alpha w_1} + \frac{(1-p)(w_2-1)}{1-\alpha+\alpha w_2} = 0.$$

This gives the formula (1) after a little algebra. Since the second derivative of $g(\alpha)$ is negative on $(0,1)$, the function $g(\alpha)$ is concave on $(0,1)$, and so $g(\alpha)$ attains its absolute maximum for the value of α in (1).

Further results for the Kelly model with a single betting object are discussed in the book Tijms (2012) and the practice paper Yoder (2021). An important issue is the volatility of your bankroll under the full Kelly strategy. That's why in practice one often uses a fractional betting fraction $c\alpha$ rather than the full betting fraction α, where c is in the range $0.3-0.5$.

4. Kelly betting with multiple betting objects

In investment situations and in sport events such as soccer matches and horse races multiple investments or bets can be simultaneously done. Imagine that opportunities to bet or invest arise at successive times $t = 1, 2, \ldots$. There are n betting objects $j = 1, \ldots, n$, where $n \geq 2$. You can simultaneously bet on one or more of these objects.

Assumption: (**a**). *At any betting opportunity, only one betting object can be successful (e.g. in a horse race only one horse can win), where object j will be successful with a given probability p_j and non-successful with probability $1 - p_j$, independently of what happened at earlier betting opportunities. Hereby*

$$0 < p_j < 1 \text{ for all } j \quad \text{and} \quad \sum_{j=1}^{n} p_j = 1.$$

(**b**). *At any betting opportunity, a stake on each non-successful object j is lost, while $f_j > 0$ dollars are added to your bankroll for every dollar staked on the successful object j. The payoffs f_j are such that $p_j f_j > 1$ for at least one object j and $\sum_{j=1}^{n} 1/f_j \geq 1$.*

The probabilities p_j are typically subjective probabilities being different for each person. For example, in horse racing you can imagine that your personal estimates of the win probability of the horses are different from the bookmaker's estimates. In the Assumption, the requirement $\sum_{j=1}^{n} p_j = 1$ can be relaxed to $\sum_{j=1}^{n} p_j \leq 1$ (introduce then an auxiliary investment object $n+1$ with f_{n+1} very close to 0 and $p_{n+1} = 1 - \sum_{i=1}^{n} p_i$).

You start with a certain bankroll V_0 and the question is how to maximize the long-run growth rate of your bankroll. The Kelly betting strategy is now characterized by parameters $\alpha_1, \ldots, \alpha_n$ such that $\alpha_i \geq 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} \alpha_i \leq 1$. Under this strategy you stake the same fraction α_i of your current bankroll in object i each time, while you keep in reserve a fraction

$$\beta = 1 - \sum_{i=1}^{n} \alpha_i$$

of your current bankroll. Denote by V_m the size of your bankroll after the mth betting opportunity and let $G_m = \frac{1}{m} \ln(V_m/V_0)$ be the growth rate of your bankroll over the first m betting opportunities. Using again the law of large numbers, a generalization of the analysis in Section 3 leads to

$$\lim_{m\to\infty} G_m = E\Big[\ln\big(\beta + \sum_{i=1}^{n} \alpha_i R_i\big)\Big] \quad \text{with probability 1,}$$

where the random vector $(R_1, \ldots, R_n)$ has the joint probability distribution

$$P(R_i = f_i,\, R_j = 0 \text{ for } j \neq i) = p_i \quad \text{for } i = 1, \ldots, n.$$

Thus the long-run growth rate of your bankroll is equal to

$$\lim_{m\to\infty} G_m = \sum_{i=1}^{n} p_i[\ln(\beta + f_i \alpha_i)] \quad \text{with probability 1.} \tag{3}$$

The goal is to find the values for the α_i's such that the long-run growth rate of your bankroll is maximal. Thus you have to solve the following optimization problem:

$$\begin{array}{lll} \text{Maximize} & f(\beta,\alpha_1,\ldots,\alpha_n) & = \sum_{i=1}^n p_i \ln(\beta + f_i\alpha_i) \\ \text{subject to} & \beta + \displaystyle\sum_{i=1}^n \alpha_i & = 1 \\ & \beta,\alpha_1,\ldots,\alpha_n & \geq 0. \end{array}$$

The objective function $f(\beta,\alpha_1,\ldots,\alpha_n)$ is concave on the convex region of feasible solutions of the optimization problem.[3] An algorithm for the optimal values of β and the α_i's can be derived from the specific Kuhn-Tucker optimality conditions for a nonlinear optimization problem with linear constraints:

$$\begin{array}{llll} \text{Maximize} & f(x_1,\ldots,x_n) & & \\ \text{subject to} & \sum_{j=1}^n a_{ij}x_j & = b_i & \text{for } i = 1,2\ldots,m, \\ & x_j & \geq 0 & \text{for } j = 1,\ldots,n. \end{array}$$

If the objective function $f(x_1,\ldots,x_n)$ is differentiable and concave on the convex set of feasible solutions of the optimization problem, the Kuhn-Tucker conditions state that an absolute maximum is attained for the feasible solution $\mathbf{x}^* = (x_1^*,\ldots,x_n^*)$ if real numbers $\lambda_1^*,\ldots,\lambda_m^*$ (Lagrangian multipliers) exist such that

$$\frac{\partial f(x^*)}{\partial x_j} - \sum_{i=1}^m \lambda_i^* a_{ij} \leq 0 \quad \text{for } j = 1,\ldots,n,$$

$$x_j^*\Big[\frac{\partial f(x^*)}{\partial x_j} - \sum_{i=1}^m \lambda_i^* a_{ij}\Big] = 0 \quad \text{for } j = 1,\ldots,n,$$

$$\sum_{j=1}^n a_{ij}x_j^* = b_i \quad \text{for } i = 1,\ldots,m,$$

$$x_j^* \geq 0 \quad \text{for } j = 1,\ldots,n.$$

A proof of this result will be outlined in the Appendix.

By the Kuhn-Tucker conditions, non-negative values $\beta,\alpha_1,\ldots,\alpha_n$ satisfying $\beta+\sum_{i=1}^n \alpha_i = 1$ provide an optimal solution for the Kelly optimization

[3] Letting $f(x,y) = \ln(x + cy)$ for variables $x,y > 0$ and constant $c > 0$, it follows that $\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - (\frac{\partial^2 f}{\partial x \partial y})^2 = 0$ and $\frac{\partial^2 f}{\partial x^2} < 0$, and so $f(x,y)$ is concave in the two variables $x,y > 0$. Using this result and the basic definition of concavity, it is now readily verified that $\sum_{i=1}^n p_i \ln(\beta + f_i\alpha_i)$ is concave on the convex set of feasible solutions.

problem if for some real number λ,

$$\frac{p_i f_i}{\beta + f_i \alpha_i} - \lambda \le 0 \ \text{ for } i = 1, \ldots, n, \quad \sum_{i=1}^{n} \frac{p_i}{\beta + f_i \alpha_i} - \lambda \le 0,$$

$$\alpha_i \Big[\frac{p_i f_i}{\beta + f_i \alpha_i} - \lambda \Big] = 0 \ \text{ for } i = 1, \ldots, n, \quad \beta \Big[\sum_{i=1}^{n} \frac{p_i}{\beta + f_i \alpha_i} - \lambda \Big] = 0.$$

The first and the third condition give

$$\frac{p_i f_i}{\beta} - \lambda \le 0 \ \text{ if } \alpha_i = 0 \quad \text{ and } \quad \frac{p_i f_i}{\beta + f_i \alpha_i} - \lambda = 0 \ \text{ if } \alpha_i > 0.$$

The key step is to assume that $\beta > 0$ in the optimal solution, that is, a positive proportion of the bankroll is always kept in reserve. This premise is reasonable in view of $p_j < 1$ for all j. Then the fourth condition becomes $\sum_{i=1}^{n} p_i/(\beta + f_i \alpha_i) = \lambda$, implying the second condition. Also, using the condition $\sum_{i=1}^{n} \alpha_i = 1 - \beta$, you easily verify that

$$\beta = \frac{1 - \sum_{i \in V} p_i/\lambda}{1 - \sum_{i \in V} 1/f_i} \quad \text{ with } V = \{ i \mid \alpha_i > 0 \}.$$

Next is matter of some manipulations to get $\lambda = 1$ and to arrive at the algorithm for the optimal values of the α_i's. We omit the further details and we suffice to give the final algorithm.

Algorithm

Step 0. Renumber the indexes such that $p_1 f_1 \ge p_2 f_2 \ge \ldots \ge p_n f_n$.

Step 1. Determine r as the largest integer k for which $\sum_{j=1}^{k} 1/f_j < 1$.

Step 2. Calculate for $k = 1, \ldots, r$ the test quantity

$$B(k) = \frac{1 - \sum_{j=1}^{k} p_j}{1 - \sum_{j=1}^{k} 1/f_j}.$$

Stop at the first index k for which $p_{k+1} f_{k+1} \le B(k)$ (and then $p_j f_j \le B(k)$ for all $j > k$). Let s be this index and let $\beta = B(s)$.

Step 3. Set $\alpha_i = p_i - \beta/f_i$ for $i = 1, \ldots, s$ and $\alpha_i = 0$ for $i > s$.

The index r satisfies $r < n$, by part (**b**) of the Assumption made. This implies that $B(s) > 0$. Therefore $\alpha_i < p_i$ for all i and so $\sum_{i=1}^{n} \alpha_i < 1$. This verifies that $\beta = 1 - \sum_{i=1}^{n} \alpha_i > 0$, that is, a positive fraction of your bankroll is kept in reserve each time. It is also noted that for $n = 1$ the algorithm boils down to the optimal betting fraction $(p_1 f_1 - 1)/(f_1 - 1)$, in agreement with the Kelly formula (2) for the case of a single betting object. Next we give two numerical examples to illustrate the algorithm.

The Kelly strategy has been developed for situations in which many betting opportunities repeat themselves under identical conditions. However, the Kelly strategy provides also a useful heuristic guideline for situations with only one betting opportunity.

Example 1. (*soccer*) Suppose that the soccer club Manchester United is hosting a match against Liverpool, and that a bookmaker is paying out 4.5 times the stake if Liverpool wins, 4.5 times the stake if the match ends in a draw, and 1.75 times the stake if Manchester United wins. You estimate Liverpool's chance of winning at 25%, the chance of the game ending in a draw at 25%, and the chance of Manchester winning at 50%. If you are prepared to bet 100 dollars, how should you bet on this match? The Kelly betting model with $n = 3$ betting objects applies, where

$$p_1 = 0.25 \text{ (win Liverpool)},\ p_2 = 0.25 \text{ (draw)},\ p_3 = 0.50 \text{ (win United)}$$
$$f_1 = f_2 = 4.5 \text{ and } f_3 = 1.75.$$

Since $p_1 f_1 = p_2 f_2 = 1.125$ and $p_3 f_3 = 0.875$, the condition $p_1 f_1 \geq p_2 f_2 \geq p_3 f_3$ is satisfied. The algorithm goes as follows:

Step 1. Since $1/f_1 = \frac{10}{45}$, $1/f_1 + 1/f_2 = \frac{20}{45}$ and $1/f_1 + 1/f_2 + 1/f_3 > 1$, the index $r = 2$.

Step 2. $B(1) = \frac{27}{28}$, $B(2) = \frac{9}{10}$ and $p_2 f_2 = 1.125 > B(1)$. This gives $s = 2$ with $\beta = B(s) = 0.9$.

Step 3. $\alpha_1 = \alpha_2 = 0.25 - \frac{0.9}{4.5} = 0.05$ and $\alpha_3 = 0$.

Thus the Kelly strategy proposes that you stake 5% of your bankroll of 100 dollars on a win for Liverpool, 5% on a draw, and 0% on a win for Manchester United. For this strategy, the subjective expected value of your bankroll after the match is $100 - 10 + 0.25 \times 4.5 \times 5 + 0.25 \times 4.5 \times 5 = 101.25$ dollars. The expected value of the percentage increase of your bankroll is 1.25%. The two concurrent bets on the soccer match act as a partial hedge for each other, reducing the overall level of risk.[4]

Example 2. (*horse race*) In a horse race there are seven horses A, B, C , D , E, F and G with respective win probabilities 40%, 25%, 20%, 7%, 4%, 3% and 1% and payoff odds 1.625:1, 2.9:1, 4.5:1, 9:1, 14:1. 17:1 and 49:1. Payoff odds a:1 means that in case of a win you will receive your stake plus a dollars for each dollar staked. Numbering the horses A, B, C, D, E, F,

[4] An interesting project would be to derive an algorithm for the case of simultaneous betting options, where the outcomes of the bets are *independent* of each other and thus more than one bet can be successful at the same time. Think of betting on a number of soccer matches that are played at the same time. How should a gambler allocate the stakes when the Kelly criterion of maximizing log-utility is used? This question is addressed in Whitrow (2007).

and G as 1 (= C), 2 (= A), 3 (= B), 4 (= D), 5 (= E), 6 (= F), and 7 (= G), the Kelly model applies with

$$p_1 = 0.2,\ p_2 = 0.4,\ p_3 = 0.25,\ p_4 = 0.07,\ p_5 = 0.04,\ p_6 = 0.03,\ p_7 = 0.01,$$
$$f_1 = 5.5,\ f_2 = 2.625,\ f_3 = 3.9, f_4 = 10,\ f_5 = 15,\ f_6 = 18,\ f_7 = 50$$

satisfying the condition of decreasing values of the $p_i f_i$'s:

$$p_1 f_1 = 1.1,\ p_2 f_2 = 1.05,\ p_3 f_3 = 0.975,\ p_4 f_4 = 0.7,$$
$$p_5 f_5 = 0.6,\ p_6 f_6 = 0.54,\ p_7 f_7 = 0.50.$$

The algorithm goes as follows:

Step 1. The index $r = 5$ is the largest value of k for which $\sum_{j=1}^{k} 1/f_j < 1$.

Step 2. $B(1) = 0.97778$, $B(2) = 0.91485$, $B(3) = 0.82956$, $B(4) = 0.98986$ and $B(5) = 2.82635$. Also, $p_2 f_2 > B(1)$, $p_3 f_3 > B(2)$, but $p_4 f_4 \leq B(3)$. This gives $s = 3$ with $\beta = B(s) = 0.82956$.

Step 3. $\alpha_1 = 0.0492$, $\alpha_2 = 0.0840$, $\alpha_3 = 0.0373$, and $\alpha_j = 0$ for $j > 3$.

Thus you bet 8.4% of your bankroll on horse A, 3.7% on horse B, 4.9% on horse C and nothing on the other horses. It is noteworthy that horse B is included in your bet, even though a bet on horse B alone is not favorable ($p_3 f_3 < 1$). The expected value of the percentage increase of your bankroll is $100 \times \sum_{j=1}^{3} (p_i f_i \alpha_i - \alpha_i) = 4.6\%$.

References

[1] Henk Tijms, *Understanding Probability*, third edition, Cambridge University Press, 2012.

[2] Nicholas Yoder, The Kelly Criterion, *Luckbox Magazine*, June 2021, https://nickyoder.com/kelly-criterion-luckbox-magazine/

[3] Chris Whitrow, Algorithms for Optimal Location of Bets on Many Simultaneous Events, *Journal of the Statistical Royal Society Series C*, Vol. 66, No 5 (2007), pp. 607-623.

Appendix: Kuhn-Tucker conditions

In this appendix an elementary derivation of the Kuhn-Tucker conditions for nonlinear programming problems is given.

The most general form of a nonlinear programming problem is

$$\begin{aligned} &\max \quad f(\mathbf{x}) \\ &\text{subject to} \quad g_i(\mathbf{x}) \leq b_i \quad \text{for } i = 1, 2, \ldots, m, \end{aligned}$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ represents the (decision) variables, m is the number of constraints for the variables and the b_i are given constants. The set of feasible solutions is defined by

$$D = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq b_i \text{ for } i = 1, 2, \ldots, m\}.$$

It is noted that the set D is convex if the $g_i(\mathbf{x})$'s are convex functions.[5] A feasible solution $\mathbf{x}^* \in D$ is said to be an *optimal solution* (global maximum) for the nonlinear programming problem if $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in D$.

For any $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$, define the *Lagrange function*

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i (g_i(\mathbf{x}) - b_i).$$

A very useful result is:

Theorem 1. *Suppose that* $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ *and* $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ *satisfy*

$$\begin{array}{ll} (i) & g_i(\mathbf{x}^*) \leq b_i \text{ for } i = 1, 2, \ldots, m \\ (ii) & \mathcal{L}(\mathbf{x}^*, \lambda^*) \geq \mathcal{L}(\mathbf{x}, \lambda^*) \text{ for all } x \in \mathbb{R}^n \\ (iii) & \lambda_i^* \geq 0 \text{ for } i = 1, 2, \ldots, m \\ (iv) & \lambda_i^* \{g_i(\mathbf{x}^*) - b_i\} = 0 \text{ for } i = 1, \ldots, m. \end{array}$$

Then $\mathbf{x}^*$ *is an optimal solution for the nonlinear programming problem.*

Proof. The proof is simple. The first condition says that $\mathbf{x}^*$ is a feasible solution. Take any other feasible solution $\mathbf{x}$. Then.

$$f(\mathbf{x}^*) \overset{\text{(iv)}}{=} f(\mathbf{x}^*) - \sum_{i=1}^{m} \lambda_i^* \Big(g_i(\mathbf{x}^*) - b_i \Big) \overset{\text{(ii)}}{\geq} f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i^* \Big(g_i(\mathbf{x}) - b_i \Big) \overset{\text{(iii)}}{\geq} f(\mathbf{x}).$$

The last inequality also uses that $g_i(\mathbf{x}) - b_i \leq 0$ for all i.

The λ_i are called the *Lagrange multipliers* and condition (iv) is called the complementary slackness condition. In fact, Theorem 1 suggests a relaxation approach in which you try to solve the difficult nonlinear programming problem by solving the unconstrained optimization problem max $\{\mathcal{L}(\mathbf{x}, \lambda) : \mathbf{x} \in \mathbb{R}^n\}$ for given non-negative values of the Lagrange multipliers, where you try to choose λ in such way the corresponding optimal solution $\mathbf{x}^*$ satisfies the conditions (i) and (iv) in Theorem 1. However, this very computationally intensive approach is not practically useful for most problems.

Suppose now that the function $f(\mathbf{x})$ is differentiable and concave, and the functions $g_i(\mathbf{x})$ for $i = 1, \ldots, m$ are differentiable and convex. Then, for any given non-negative Lagrange multipliers λ_i, the Lagrange function $\mathcal{L}(\mathbf{x}, \lambda)$ is concave as function of $\mathbf{x}$, as is easily verified by noting that a function $-h$ is concave if h is convex and that a finite sum of concave functions is also concave. In this case condition (ii) is equivalent with

$$\nabla_x \mathcal{L}(\mathbf{x}^*, \lambda) = 0.$$

[5] A nice treatment of convexity and concavity of functions of several variables can be found in https://mjo.osborne.economics.utoronto.ca/index.php/tutorial/index/1/cvn/t

As consequence of Theorem 1, we now obtain the following important main theorem in nonlinear programming:

Theorem 2. *Suppose that the function $f(\mathbf{x})$ is differentiable and concave, and the functions $g_i(\mathbf{x})$ are differentiable and convex for $i = 1, \ldots, m$. If $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ and $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ satisfy*

$$(i) \quad g_i(\mathbf{x}^*) \leq b_i \quad \text{for } i = 1, 2, \ldots, m$$

$$(ii) \quad \frac{\partial f(\mathbf{x}^*)}{\partial x_j} - \sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} = 0 \quad \text{for } j = 1, \ldots, n$$

$$(iii) \quad \lambda_i^* \geq 0 \quad \text{for } i = 1, 2, \ldots, m$$

$$(iv) \quad \lambda_i^* \{g_i(\mathbf{x}^*) - b_i\} = 0 \quad \text{for } i = 1, \ldots, m.$$

Then $\mathbf{x}^*$ is an optimal solution for the nonlinear programming problem.

The four conditions in Theorem 2 are called *Kuhn-Tucker conditions*. These conditions generalize the optimality conditions for linear programming, where the λ_i's play the role of the dual variables. The Kuhn-Tucker conditions stated in Section 4 for the nonlinear program with linear constraints are rather easily derived from Theorem 2. To do so, note that linear functions are convex, replace each non-negativity requirement $x_j \geq 0$ by the inequality $g_{m+j}(\mathbf{x}) \leq 0$ with $g_{m+j}(\mathbf{x}) = -x_j$, and replace each linear equality $\sum_{j=1}^{n} a_{ij} x_j = b_i$ by the two inequalities $\sum_{j=1}^{n} a_{ij} x_j \leq b_i$ and $\sum_{j=1}^{n} -a_{ij} x_j \leq -b_i$. The technical details of the derivation are left to the reader.

Remark. The Kuhn-Tucker conditions are the very foundation for several computational methods for solving nonlinear programming problems, such as the quadratic programming problem with a quadratic criterion function and linear constraints. The Kuhn-Tucker conditions for quadratic programming problems have a simple form that can make solutions considerably easier to obtain than for general linear programming problems. Choosing an algorithm for a nonlinear programming problem is often difficult because no one algorithm can be expected to work for every kind of nonlinear programming problems. Software platforms as AIMMS, AMPL, GAMS, Python and **R** include a number of optimizers for nonlinear programming problems with the hope that one of these methods will suffice for the given problem. It should be pointed out that in many problems it is difficult to determine whether the objective function is concave in the feasible region and hence it is difficult to guarantee convergence to a global optimum.