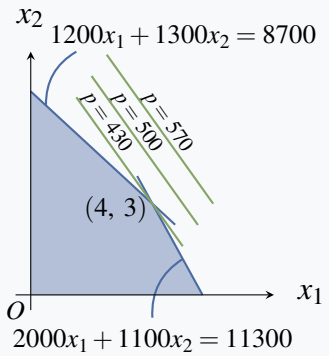
Example 8.10.3



A manufacturer makes x_1 units of product 1, and x_2 units of product 2, at a profit of \$70 and \$50 per unit respectively, and wants to choose x_1 and x_2 to maximize the total profit $p(x_1, x_2) = 70x_1 + 50x_2$. However x_1 and x_2 are not arbitrary; for example, $x_1 \geq 0$ and $x_2 \geq 0$. Other conditions also come into play. Each unit of product 1 costs \$1200 to produce and requires 2000 square feet of warehouse space; each unit of product 2 costs \$1300 to produce and requires 1100 square feet of space. If the total warehouse space is 11 300 square feet, and if the total production budget is \$8700, x_1 and x_2 must also satisfy the conditions

$$2000x_1 + 1100x_2 \leq 11300$$
$$1200x_1 + 1300x_2 \leq 8700$$

The feasible region in the plane satisfying these constraints (and $x_1 \geq 0$, $x_2 \geq 0$) is shaded in the diagram. If the profit equation $70x_1 + 50x_2 = p$ is plotted for various values of p, the resulting lines are parallel, with p increasing with distance from the origin. Hence the best choice occurs for the line $70x_1 + 50x_2 = 430$ that touches the shaded region at the point $(4, 3)$. So the profit p has a maximum of $p = 430$ for $x_1 = 4$ units and $x_2 = 3$ units.

Example 8.10.3 is a simple case of the general **linear programming** problem[23] which arises in economic, management, network, and scheduling applications. Here the objective function is a linear combination $q = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ of the variables, and the feasible region consists of the vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ in $\mathbb{R}^n$ which satisfy a set of linear inequalities of the form $b_1x_1 + b_2x_2 + \cdots + b_nx_n \leq b$. There is a good method (an extension of the gaussian algorithm) called the **simplex algorithm** for finding the maximum and minimum values of q when $\mathbf{x}$ ranges over such a feasible set. As Example 8.10.3 suggests, the optimal values turn out to be vertices of the feasible set. In particular, they are on the boundary of the feasible region, as is the case in Theorem 8.10.1.

8.11 An Application to Statistical Principal Component Analysis

Linear algebra is important in multivariate analysis in statistics, and we conclude with a very short look at one application of diagonalization in this area. A main feature of probability and statistics is the idea of a *random variable* X, that is a real-valued function which takes its values according to a probability law (called its *distribution*). Random variables occur in a wide variety of contexts; examples include the number of meteors falling per square kilometre in a given region, the price of a share of a stock, or the duration of a long distance telephone call from a certain city.

The values of a random variable X are distributed about a central number μ, called the *mean* of X. The mean can be calculated from the distribution as the *expectation* $E(X) = \mu$ of the random variable X.

[23] More information is available in "Linear Programming and Extensions" by N. Wu and R. Coppins, McGraw-Hill, 1981.

Functions of a random variable are again random variables. In particular, $(X-\mu)^2$ is a random variable, and the *variance* of the random variable X, denoted $\text{var}(X)$, is defined to be the number

$$\text{var}(X) = E\{(X-\mu)^2\} \quad \text{where } \mu = E(X)$$

It is not difficult to see that $\text{var}(X) \geq 0$ for every random variable X. The number $\sigma = \sqrt{\text{var}(X)}$ is called the *standard deviation* of X, and is a measure of how much the values of X are spread about the mean μ of X. A main goal of statistical inference is finding reliable methods for estimating the mean and the standard deviation of a random variable X by sampling the values of X.

If two random variables X and Y are given, and their joint distribution is known, then functions of X and Y are also random variables. In particular, $X+Y$ and aX are random variables for any real number a, and we have

$$E(X+Y) = E(X)+E(Y) \quad \text{and} \quad E(aX) = aE(X).^{24}$$

An important question is how much the random variables X and Y depend on each other. One measure of this is the *covariance* of X and Y, denoted $\text{cov}(X, Y)$, defined by

$$\text{cov}(X, Y) = E\{(X-\mu)(Y-\upsilon)\} \quad \text{where } \mu = E(X) \text{ and } \upsilon = E(Y)$$

Clearly, $\text{cov}(X, X) = \text{var}(X)$. If $\text{cov}(X, Y) = 0$ then X and Y have little relationship to each other and are said to be *uncorrelated*.[25]

Multivariate statistical analysis deals with a family $X_1, X_2, \ldots, X_n$ of random variables with means $\mu_i = E(X_i)$ and variances $\sigma_i^2 = \text{var}(X_i)$ for each i. Let $\sigma_{ij} = \text{cov}(X_i, X_j)$ denote the covariance of X_i and X_j. Then the *covariance matrix* of the random variables $X_1, X_2, \ldots, X_n$ is defined to be the $n \times n$ matrix

$$\Sigma = [\sigma_{ij}]$$

whose (i, j)-entry is σ_{ij}. The matrix Σ is clearly symmetric; in fact it can be shown that Σ is **positive semidefinite** in the sense that $\lambda \geq 0$ for every eigenvalue λ of Σ. (In reality, Σ is positive definite in most cases of interest.) So suppose that the eigenvalues of Σ are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. The principal axes theorem (Theorem 8.2.2) shows that an orthogonal matrix P exists such that

$$P^T \Sigma P = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

If we write $\overline{X} = (X_1, X_2, \ldots, X_n)$, the procedure for diagonalizing a quadratic form gives new variables $\overline{Y} = (Y_1, Y_2, \ldots, Y_n)$ defined by

$$\overline{Y} = P^T \overline{X}$$

These new random variables $Y_1, Y_2, \ldots, Y_n$ are called the **principal components** of the original random variables X_i, and are linear combinations of the X_i. Furthermore, it can be shown that

$$\text{cov}(Y_i, Y_j) = 0 \text{ if } i \neq j \quad \text{and} \quad \text{var}(Y_i) = \lambda_i \quad \text{for each } i$$

Of course the principal components Y_i point along the principal axes of the quadratic form $q = \overline{X}^T \Sigma \overline{X}$.

[24] Hence $E(\ )$ is a linear transformation from the vector space of all random variables to the space of real numbers.

[25] If X and Y are independent in the sense of probability theory, then they are uncorrelated; however, the converse is not true in general.

The sum of the variances of a set of random variables is called the **total variance** of the variables, and determining the source of this total variance is one of the benefits of principal component analysis. The fact that the matrices Σ and $\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are similar means that they have the same trace, that is,

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{nn} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

This means that the principal components Y_i have the same total variance as the original random variables X_i. Moreover, the fact that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ means that most of this variance resides in the first few Y_i. In practice, statisticians find that studying these first few Y_i (and ignoring the rest) gives an accurate analysis of the total system variability. This results in substantial data reduction since often only a few Y_i suffice for all practical purposes. Furthermore, these Y_i are easily obtained as linear combinations of the X_i. Finally, the analysis of the principal components often reveals relationships among the X_i that were not previously suspected, and so results in interpretations that would not otherwise have been made.