6. GAMES AND DUALITY

In trying to understand how Lagrange multiplier vectors $\bar{y}$ might be generated or utilized in computational schemes, the case of linear programming is instructive. Inspection of the Kuhn-Tucker conditions in that case reveals that these vectors solve a mysterious problem of optimization inextricably tied to the given one. Pursuing the mystery further, we are led to an interesting branch of modern mathematics: game theory. We'll look briefly at this theory and use it to develop the fact that for convex programming problems quite generally the Lagrange multiplier vectors associated with optimality can be obtained in principle by solving an auxiliary problem said to be "dual" to the given "primal" problem.

Kuhn-Tucker conditions in linear programming: Consider $(\mathcal{P})$ in the special case of a linear programming problem in so-called *primal canonical format*:

$$(\mathcal{P}_{\rm lin}) \qquad \begin{aligned} &\text{minimize } \textstyle\sum_{j=1}^n c_j x_j \text{ subject to} \\ &\textstyle\sum_{j=1}^n a_{ij} x_j \geq b_i \text{ for } i = 1, \ldots, m, \\ &\qquad\qquad x_j \geq 0 \text{ for } j = 1, \ldots, n, \end{aligned}$$

which corresponds in the conventional format to taking $f_0(x) := \sum_{j=1}^n c_j x_j$ with $f_i(x) = b_i - \sum_{j=1}^n a_{ij} x_j$ for $i = 1, \ldots, m$ and $X = I\!R^n_+$. Because all the constraint functions correspond to inequalities, we have $Y = I\!R^m_+$. The Lagrangian is

$$\begin{aligned} L(x,y) &= \textstyle\sum_{j=1}^n c_j x_j + \sum_{i=1}^m y_i \Big(b_i - \sum_{j=1}^n a_{ij} x_j \Big) \\ &= \textstyle\sum_{i=1}^m b_i y_i + \sum_{j=1}^n x_j \Big(c_j - \sum_{i=1}^m y_i a_{ij} \Big) \text{ on } I\!R^n_+ \times I\!R^m_+. \end{aligned}$$

In recalling the nature of normal cones to the orthants $I\!R^n_+$ and $I\!R^m_+$, we see that the Kuhn-Tucker conditions in this case come out as

$$\begin{cases} \bar{x}_j \geq 0, \ \Big(c_j - \sum_{i=1}^m \bar{y}_i a_{ij} \Big) \geq 0, \ \bar{x}_j \Big(c_j - \sum_{i=1}^m \bar{y}_i a_{ij} \Big) = 0 & \text{for } j = 1, \ldots, n, \\ \bar{y}_i \geq 0, \ \Big(\sum_{j=1}^n a_{ij} \bar{x}_j - b_i \Big) \geq 0, \ \bar{y}_i \Big(\sum_{j=1}^n a_{ij} \bar{x}_j - b_i \Big) = 0 & \text{for } i = 1, \ldots, m. \end{cases}$$

Complementary slackness: These relationships between $\bar{x}$ and $\bar{y}$ are known as the *complementary slackness conditions* in linear programming. They list for each index j or i a pair of inequalities (one on $\bar{x}$ and one on $\bar{y}$), requiring that at most one of the two can be "slack," i.e., satisfied with strict inequality. By Theorems 11 and 12, $\bar{x}$ is optimal in $(\mathcal{P}_{\rm lin})$ if and only if this holds for some $\bar{y}$.

Dual linear programming problem: The symmetry in these conditions is tantalizing. There turns out to be a connection with the following problem, said to be *dual* to $(\mathcal{P}_{\text{lin}})$, which is a linear programming problem in so-called *dual canonical format*:

$$(\mathcal{D}_{\text{lin}}) \qquad \begin{aligned} &\text{maximize } \sum\nolimits_{i=1}^{m} b_i y_i \text{ subject to} \\ &\sum\nolimits_{i=1}^{m} y_i a_{ij} \le c_j \text{ for } j = 1, \dots, n, \\ &\qquad y_i \ge 0 \text{ for } i = 1, \dots, m. \end{aligned}$$

To see the formal relationship between $(\mathcal{D}_{\text{lin}})$ and $(\mathcal{P}_{\text{lin}})$, begin by converting $(\mathcal{D}_{\text{lin}})$ from dual canonical format to primal canonical format:

$$\begin{aligned} &\text{minimize } \sum\nolimits_{i=1}^{m} [-b_i] y_i \text{ subject to} \\ &\sum\nolimits_{i=1}^{m} y_i [-a_{ij}] \ge [-c_j] \text{ for } j = 1, \dots, n, \\ &\qquad y_i \ge 0 \text{ for } i = 1, \dots, m. \end{aligned}$$

As an act of faith, permit the symbols $\bar{x}_j$ to be used for the Lagrange multipliers associated with optimality in this problem, without presupposing for now any tie with the use of the notation $\bar{x}$ above. The Kuhn-Tucker conditions for this problem in their complementary slackness formulation emerge then as

$$\begin{cases} \bar{y}_i \ge 0, \ \left([-b_i] - \sum\nolimits_{j=1}^{n} [-a_{ij}] \bar{x}_j\right) \ge 0, \ \bar{y}_i \left([-b_i] - \sum\nolimits_{j=1}^{n} [-a_{ij}] \bar{x}_j\right) = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } i = 1, \dots, m, \\ \bar{x}_j \ge 0, \ \left(\sum\nolimits_{i=1}^{m} \bar{y}_i [-a_{ij}] - [-c_j]\right) \ge 0, \ \bar{x}_j \left(\sum\nolimits_{i=1}^{m} \bar{y}_i [-a_{ij}] - [-c_j]\right) = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } j = 1, \dots, n. \end{cases}$$

But these are identical to the complementary slackness conditions we had before. Problems $(\mathcal{P}_{\text{lin}})$ and $(\mathcal{D}_{\text{lin}})$ thus turn out to share the very same optimality conditions! Neither can be solved without somehow, explicitly or implicitly, solving the other as well. This has important consequences.

Symmetry: The reason for treating $(\mathcal{D}_{\text{lin}})$ as a problem of maximization instead of minimization is to bring out not only the symmetry in this switch of signs, but to promote a relationship of optimal values which is given in the theorem below. It's interesting to note that after $(\mathcal{D}_{\text{lin}})$ has been converted to a problem in primal canonical form—denoted say by $(\mathcal{P}'_{\text{lin}})$—it will in turn have an associated dual problem—say $(\mathcal{D}'_{\text{lin}})$. But this can be seen to be none other than the problem obtained by converting $(\mathcal{P}_{\text{lin}})$ from primal canonical form to dual canonical form. Thus, the relationship between the two problems is completely balanced: each can be viewed as dual to the other.

THEOREM 14 (duality in linear programming). *For a linear programming problem $(\mathcal{P}_{\rm lin})$ in primal canonical form and the associated linear programming problem $(\mathcal{D}_{\rm lin})$ in dual canonical form, the following properties of a pair of vectors $\bar{x}$ and $\bar{y}$ are equivalent:*

(a) *$\bar{x}$ is an optimal solution to $(\mathcal{P}_{\rm lin})$, and $\bar{y}$ is an associated Lagrange multiplier vector in the Kuhn-Tucker conditions for $(\mathcal{P}_{\rm lin})$ at $\bar{x}$;*

(b) *$\bar{y}$ is an optimal solution to $(\mathcal{D}_{\rm lin})$, and $\bar{x}$ is an associated Lagrange multiplier vector in the Kuhn-Tucker conditions for $(\mathcal{D}_{\rm lin})$ at $\bar{y}$;*

(c) *$\bar{x}$ and $\bar{y}$ are optimal solutions to $(\mathcal{P}_{\rm lin})$ and $(\mathcal{D}_{\rm lin})$, respectively;*

(d) *$\bar{x}$ is a feasible solution to $(\mathcal{P}_{\rm lin})$, $\bar{y}$ is a feasible solution to $(\mathcal{D}_{\rm lin})$, and the objective function values at these points are equal: $\sum_{j=1}^n c_j\bar{x}_j = \sum_{i=1}^m b_i\bar{y}_i$;*

(e) *$\bar{x}$ and $\bar{y}$ satisfy the complementary slackness conditions*

$$\begin{cases} \bar{x}_j \geq 0, \ \left(c_j - \sum_{i=1}^m \bar{y}_i a_{ij}\right) \geq 0, \ \bar{x}_j\left(c_j - \sum_{i=1}^m \bar{y}_i a_{ij}\right) = 0 & \text{for } j = 1,\dots,n, \\ \bar{y}_i \geq 0, \ \left(\sum_{j=1}^n a_{ij}\bar{x}_j - b_i\right) \geq 0, \ \bar{y}_i\left(\sum_{j=1}^n a_{ij}\bar{x}_j - b_i\right) = 0 & \text{for } i = 1,\dots,m. \end{cases}$$

Furthermore, if either $(\mathcal{P}_{\rm lin})$ or $(\mathcal{D}_{\rm lin})$ has an optimal solution, then both have optimal solutions, and

$$\left[\text{optimal value in } \left(\mathcal{P}_{\rm lin}\right)\right] = \left[\text{optimal value in } \left(\mathcal{D}_{\rm lin}\right)\right].$$

Proof. The equivalence of (a) and (b) with (e) has been ascertained in the preceding discussion. We also know through Theorems 11 and 12 that the optimality of $\bar{x}$ in $(\mathcal{P}_{\rm lin})$ is equivalent to the existence of a vector $\bar{y}$ satisfying these conditions, and likewise that the optimality of $\bar{y}$ in $(\mathcal{D}_{\rm lin})$ is equivalent to the existence of a vector $\bar{x}$ satisfying these conditions. Hence if either problem has an optimal solution, the other must have one as well. Observe next that the complementary slackness conditions in (e) entail

$$\sum_{j=1}^n c_j\bar{x}_j - \sum_{i=1}^m \bar{y}_i b_i = \sum_{j=1}^n \Big[c_j - \sum_{i=1}^m \bar{y}_i a_{ij}\Big]\bar{x}_j + \sum_{i=1}^m \bar{y}_i\Big[\sum_{j=1}^n a_{ij}\bar{x}_j - b_i\Big] = 0.$$

It follows that when optimal solutions exist the optimal values in the two problems coincide. This leads to the further equivalence of (a)-(b)-(e) with (c) and with (d). □

Existence of optimal solutions in linear programming: Incidentally, a special fact in linear—and also quadratic—programming, which however we won't work at establishing here, is that if the optimal value in such a problem is finite then an optimal solution must exist. This in contrast to the situation for more general instances of problem $(\mathcal{P})$, where the optimal value might be finite but only approachable asymptotically through a sequence of feasible solutions x^ν with $|x^\nu| \to \infty$. For such problems, Theorem 1 governs the existence of optimal solutions, but in linear and quadratic programming the power of Theorem 1 isn't needed.

Existence through feasibility: It follows from the fact about linear programming just mentioned that, for a primal-dual pair of problems $(\mathcal{P}_{\rm lin})$ and $(\mathcal{D}_{\rm lin})$, if feasible solutions exist to both problems then both also have optimal solutions.

Argument: In a slight extension of the proof of Theorem 14, we observe that for any feasible solutions to the two problems we have

$$\sum_{j=1}^n c_j x_j - \sum_{i=1}^m y_i b_i \;=\; \sum_{j=1}^n \Big[c_j - \sum_{i=1}^m y_i a_{ij}\Big] x_j + \sum_{i=1}^m y_i \Big[\sum_{j=1}^n a_{ij} x_j - b_i\Big] \;\geq\; 0.$$

Then the value $\sum_{i=1}^m y_i b_i$ is a lower bound to the optimal value in $(\mathcal{P}_{\rm lin})$ (implying it must be finite), while at the same time $\sum_{j=1}^n c_j x_j$ is an upper bound to the optimal value in $(\mathcal{D}_{\rm lin})$ (implying that it too must be finite).

Dantzig's simplex method in linear programming: The complementary slackness conditions are the algebraic foundation for an important numerical technique for solving linear programming problems, the very first, which actually was the breakthrough that got optimization rolling around 1950 as a modern subject with impressive practical applications. In theory, the task of finding an optimal solution $\bar{x}$ to $(\mathcal{P}_{\rm lin})$ is equivalent to that of finding a pair of vectors $\bar{x}$ and $\bar{y}$ for which these conditions hold. Trying directly to come up with solutions to systems of linear inequalities is a daunting challenge, but solving linear equations is more attractive, and this can be made the focus through the fact that the complementary slackness conditions require at least one of the inequalities for each index j or i to hold as an equation. Let's approach this in terms of selecting of two index sets $I \subset \{i = 1, \ldots, m\}$ and $J \subset \{j = 1, \ldots, n\}$, and associating with them following system of $n + m$ equations in the $n + m$ unknowns $\bar{x}_j$ and $\bar{y}_i$:

$$(I, J) \qquad \begin{array}{ll} \sum_{j=1}^m a_{ij}\bar{x}_j - b_i = 0 \text{ for } i \in I, & \bar{y}_i = 0 \text{ for } i \notin I, \\ c_j - \sum_{i=1}^m \bar{y}_i a_{ij} = 0 \text{ for } j \in J, & \bar{x}_j = 0 \text{ for } j \notin J. \end{array}$$

To say that $\bar{x}$ and $\bar{y}$ satisfy the complementary slackness conditions is to say that for *some* choice of I and J, this (I, J)-system of linear equations will have a solution $(\bar{x}, \bar{y})$ for which the inequalities

$$\begin{array}{ll} \sum_{j=1}^m a_{ij}\bar{x}_j - b_i \geq 0 \text{ for } i \notin I, & \bar{y}_i \geq 0 \text{ for } i \in I, \\ c_j - \sum_{i=1}^m \bar{y}_i a_{ij} \geq 0 \text{ for } j \notin J, & \bar{x}_j \geq 0 \text{ for } j \in J \end{array}$$

happen to be satisfied as well. The main fact is that there is only a *finite collection* of (I, J)-systems, because there are only finitely many ways of choosing I and J. Moreover, it can be shown that one can limit attention to (I, J)-systems that are nondegenerate, i.e., have nonsingular matrix so that there's a unique corresponding pair $(\bar{x}, \bar{y})$. The prospect is thereby raised of searching by computer through a finite list of possible candidate pairs $(\bar{x}, \bar{y})$, each obtained by solving a certain system of linear equations, checking each time whether the desired inequalities are satisfied too, until a pair is found that meets the test of the complementary slackness conditions and thereby provides an optimal solution to the linear programming problem $(\mathcal{P}_{\text{lin}})$.

Of course, even with this idea one is far from a practical method of computation, because the number of (I, J)-systems that would have to be inspected is likely to be awesomely astronomical, far beyond the capability of thousands of the fastest computers laboring for thousands of years! But fortunately it's not necessary to look at all such systems. There are ways of starting with one such (I, J)-system and then modifying the choice of I and J in tiny steps in such a manner that "improvements" are continually made. We won't go into the scheme further here, but this is the contribution of Dantzig that so changed the world of optimization at the beginning of the computer age. Nowadays there are other ways of solving linear programming problems, but Dantzig's so-called "simplex method" is still competitive and used.

Duality more generally: The facts about the tight relationship between the linear programming problems $(\mathcal{P}_{\text{lin}})$ and $(\mathcal{D}_{\text{lin}})$ raise more questions than they answer. What is the "explanation" for this phenomenon, and what significance, not just technical, can be ascribed to the Lagrange multiplier values that appear? How, for instance, might they be interpreted in a particular application? These issues go far beyond linear programming. To lay the groundwork for their analysis, we need to spend some time with elementary game theory.

Lagrangian framework as the key: Duality in linear programming, in its manifestation so far, has grown out of complementary slackness form of the Kuhn-Tucker conditions that characterize optimality. As we move to from linear programming to greater generality, we'll be focusing more on the Lagrangian form of the Kuhn-Tucker conditions and on the Lagrangian L itself as a special function on a certain product set $X \times Y$. Cases where a pair $(\bar{x}, \bar{y})$ constitutes a "Lagrangian saddle point" will be especially important. For the time being, though, we'll profit from allowing the L, X, Y, notation to be used in a broader context, where they need not signify anything about a Lagrangian.

Two-person, zero-sum games: Consider *any* function $L : X \times Y \to I\!R$ for *any* nonempty sets X and Y, not necessarily even in $I\!R^n$ and $I\!R^n$. There's an associated game, played by two agents, called Player 1 and Player 2. Player 1 selects some $x \in X$, while Player 2 selects some $y \in Y$. The choices are revealed simultaneously. Then Player 1 must pay $\$L(x, y)$ to Player 2—that's all there is to it.

Direction of payment: Because no restriction has been placed on the sign of $L(x, y)$, the game is *not* necessarily loaded against Player 1 in favor of Player 2. The payment of a negative amount $\$L(x, y)$ from Player 1 to Player 2 is code for money actually flowing in the opposite direction.)

Terminology: In general, X and Y are called the "strategy sets" for Players 1 and 2, while L is the "payoff function."

Generality of the game concept: This abstract model of a game may appear too special to be worthy of the name. Yet appearances can be deceiving: games such as *chess* and even *poker* can in principle be covered. We won't be concerned with those games here, but it's worth sketching them into the picture anyway.

Chess as a two-person, zero-sum game. In chess, each element x of the set X at the disposal of Player 1, with the white pieces, is a particular policy which specifies encyclopedically what Player 1 should do within all possible circumstances that might arise. For instance, just one tiny part of such a policy x would be a prescription like "if after 17 moves the arrangement of the pieces on the board is such-and-such, and the history of play that brought it about is such-and-such, then such-and-such move should be made." The elements $y \in Y$ have similar meaning as policies for Player 2, with the black pieces. Clearly, in choosing such policies independently the two players are merely deciding in advance how they will respond to whatever may unfold as the chess game is played. From (x, y) the outcome is unambiguously determined: checkmate for Player 1, checkmate for Player 2, or a draw. Define $L(x, y)$ to be -1, 1 or 0 according to these three cases. The chess game is represented then by X, Y, and L. It's *not* claimed that this is a *practical* representation, because the sets X and Y are impossibly large, but conceptually it's correct.

Poker as a two-person, zero-sum game. The game of poker can be handled in similar fashion. It's payoffs are probabilistic "expectations," since they depend to a certain extent on random events beyond the players' control. Nonetheless, features of poker like bluffing can be captured in the model and evaluated for their effects. Poker for more than two players can be treated as an N-person game.

Saddle points: The term "saddle point" has various usages in mathematics, but in game theory and optimization it always refers to the following concept, which in the particular context of a two-person zero-sum game expresses a kind of solution to the conflict between the two players. A *saddle point* of a function L over a product set $X \times Y$ is a pair $(\bar{x}, \bar{y}) \in X \times Y$ such that

$$L(x, \bar{y}) \geq L(\bar{x}, \bar{y}) \geq L(\bar{x}, y) \text{ for all } x \in X \text{ and } y \in Y.$$

This means that the minimum of $L(x, \bar{y})$ with respect to $x \in X$ is attained at $\bar{x}$, whereas the maximum of $L(\bar{x}, y)$ with respect to $y \in Y$ is attained at $\bar{y}$.

Equilibrium interpretation: In the setting of a game with strategy sets X and Y and payoff function L, a saddle point $(\bar{x}, \bar{y})$ of L over $X \times Y$ captures a situation in which neither player, acting unilaterally, has any incentive for deviating from the strategies $\bar{x}$ and $\bar{y}$. In selecting $\bar{x}$, Player 1 can guarantee that the amount paid to Player 2 won't exceed $\$L(\bar{x}, \bar{y})$, even if Player 2 were aware in advance that $\bar{x}$ would be chosen. This results from the fact that $L(\bar{x}, \bar{y}) \geq L(\bar{x}, y)$ for all $y \in Y$ (as half of the defining property of a "saddle point"). At the same time, in selecting $\bar{y}$, Player 2 can guarantee that the amount received from Player 1 won't fall short of $\$L(\bar{x}, \bar{y})$, regardless of whether Player 1 acts with knowledge of this choice or not. This is because $L(\bar{x}, \bar{y}) \leq L(x, \bar{y})$ for all $x \in X$.

Fairness: The game is *fair* if the equilibrium payoff amount $\$L(\bar{x}, \bar{y})$ at a saddle point $(\bar{x}, \bar{y})$ is 0. Then neither player has an advantage over the other. Of course the issue of whether a saddle point exists at all for a particular case of X, Y and L is separate. Without crucial assumptions being satisfied by L, X and Y, there might not be one. The existence of saddle points is studied in *minimax theory*, but we'll be looking at some special cases.

Chess and poker: No one knows whether a saddle point exists for chess in the game formulation we've given. What's known is that if "randomized" policies involving probabilistic play are introduced in a certain way, then a saddle point *does* exist. ("Randomized" policies allow for prescribing in a each circumstance not necessarily a fixed move, but a probability distribution among several moves, for example two possible moves with a 50/50 distribution, the ultimate choice to be resolved at the last moment by flipping a coin. Correspondingly, the payoffs are modified to reflect *chances* of winning.) Whether the equilibrium value $L(\bar{x}, \bar{y})$ in such a model is 0, testifying that chess is a fair game when randomized policies are admitted, is unknown. The theory of poker is in a similar state.

Gradient condition associated with a saddle point: In the important case of a game in which X and Y happen to be closed subsets of $I\!R^n$ and $I\!R^m$ while L is a function of class $\mathcal{C}^1$, our work on constrained optimization gives some insights into a saddle point. In that case a first-order necessary condition for $(\bar{x}, \bar{y})$ to be a saddle point of L over $X \times Y$ is that

$$-\nabla_x L(\bar{x}, \bar{y}) \in N_X(\bar{x}), \qquad \nabla_y L(\bar{x}, \bar{y}) \in N_Y(\bar{x}).$$

This follows from applying Theorem 9(a) first to the minimization of $L(x, \bar{y})$ in $x \in X$ and then to the maximization of $L(\bar{x}, y)$ in $y \in Y$. By Theorem 9(b) we see that, conversely, this gradient condition is sufficient for a saddle point in situations where $L(x, y)$ is convex in x for each $y \in Y$, and on the other hand concave in y for each $x \in X$.

Preview about Lagrangians: This gradient condition turns into the Lagrangian form of the Kuhn-Tucker conditions when L is the Lagrangian associated with an optimization problem $(\mathcal{P})$ in conventional format. This is the pipeline down which we are eventually headed.

Broader game models: In an N*-person game* Player k for $k = 1, \ldots, N$ selects an element $x_k \in X_k$, and the choices are revealed simultaneously. Then Player k must pay the (positive, zero, or negative) amount $L_k(x_1, \ldots, x_N)$—to the "great accountant in the sky," or whatever. The game is *zero-sum* if

$$\sum_{k=1}^{N} L_k(x_1, \ldots, x_N) = 0.$$

The interpretation can then be made that all the payments for positive L_k values go into a single pot and get redistributed in accordance with the negative L_k values that are present. In the two-person, zero-sum case one simply has $L_1(x_1, x_2) = -L_2(x_1, x_2)$, and the notation for this quantity can be shifted to $L(x, y)$.

For games that aren't zero-sum, no actual exchange of anything needs to be envisioned, and there is no reason to focus on a medium of exchange, such as money. The units in which the values of L_k are measured can be personal to Player k and different from those for the other players. (A popular choice is for L_k, or rather its negative through a switch to maximization, to be a so-called *utility* function for Player k, for the construction of which a large body of theory is available.) This makes it possible to use game theory in the modeling of social and economic situations that don't necessarily reduce to competition alone. For instance, *cooperative* games can be studied, where everyone can win if the actions are properly coordinated. Theory must then address the mechanisms of achieving coordination among players.

The central feature of any N-person game is that the consequences for Player k of choosing x_k depend not only on x_k itself, over which Player k has control, but also on the decisions made by the other players.

Nash equilibrium: A basic concept of "solution" in the theory of N-person games is that of *Nash equilibrium*. This refers to having $(\bar{x}_1, \ldots, \bar{x}_N) \in X_1 \times \cdots \times X_N$ such that Player 1 faces

$$L_1(x_1, \bar{x}_2, \ldots, \bar{x}_N) \geq L_1(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N) \text{ for all } x_1 \in X_1,$$

while Player 2 faces

$$L_2(\bar{x}_1, x_2, \ldots, \bar{x}_N) \geq L_2(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N) \text{ for all } x_2 \in X_2,$$

and similarly for Player 3 to Player N (when present). As with a saddle point, the interpretation is that no single player would have incentive for *unilaterally* making a different decision. For a two-person zero-sum game, Nash equilibrium reduces to a saddle point. Other concepts of equilibrium have also been explored theoretically, like ones involving the formation of "coalitions" among the players, reinforced perhaps by side payments to keep members in line, these side payments coming out of the proceeds obtained by the coalition.

Game theory nowadays: The greatest interest in game theory has come from mathematical economists, who have developed and used it in many ways. In fact the subject was invented by John von Neumann, one of the most remarkable mathematicians of the 20th century, especially for that purpose. Hopes that game theory might lead to major advances in the study of human behavior haven't been borne out, however. But at this moment interest on the level of business applications is resurgent.

Optimization problems derived from a game: Associated with any two-person, zero-sum game specified by a general choice of X, Y and L, there are two complementary problems of optimization. The study of these problems provides further insight into the role of saddle points and, eventually, the ways that Lagrange multiplier vectors can be determined, interpreted and utilized. The problems in question result from adopting a very conservative approach to playing the game, amounting to little more than a worst-case analysis. Whether or not one would really be content with playing the game in this manner in practice, the approach does lead to impressive mathematical results.

Minimax strategy problem for Player 1: To determine $\bar{x}$, Player 1 should solve

$$(\mathcal{P}_1) \qquad \text{minimize } f(x) \text{ over all } x \in X, \text{ where } f(x) := \sup_{y \in Y} L(x,y).$$

In other words, for each $x \in X$, Player 1 should look at the value $f(x)$, which indicates the worst that could possibly happen if x were the element selected (the measure being in terms of how high a payment might have to be made to Player 2, in the absence of any way to predict what Player 2 will actually do). The choice of $\bar{x}$ should be made to ameliorate this as far as possible.

Minimax strategy problem for Player 2: To determine $\bar{y}$, Player 2 should solve

$$(\mathcal{P}_2) \qquad \text{maximize } g(y) \text{ over all } y \in Y, \text{ where } g(y) := \inf_{x \in X} L(x,y).$$

In other words, for each $y \in Y$, Player 2 should look at the value $g(y)$, which indicates the worst that could possibly happen if y were the element selected (the measure being in terms of how low a payment might be forthcoming from Player 1, in the absence of any way to predict what Player 1 will actually do). The choice of $\bar{y}$ should be made to ameliorate this as far as possible.

Fundamental relation in optimal values: Let $\bar{\alpha}$ denote the optimal value in Player 1's problem and $\bar{\beta}$ the optimal value in Player 2's problem. Then

$$\bar{\alpha} \geq \bar{\beta}.$$

This is evident on intuitive grounds, even without assurance that optimal solutions exist to the two problems. For any $\alpha > \bar{\alpha}$, Player 1 can choose $x \in X$ with $f(x) \leq \alpha$ and thereby be certain of not having to pay more than the amount α. On the other hand, for any $\beta < \bar{\beta}$, Player 2 can choose $y \in Y$ with $g(y) \geq \beta$ and be certain of getting at least the amount β from Player 1. Seen more fully, the definition of f gives $f(x) \geq L(x,y)$ for any $y \in Y$, while the definition of g gives $g(y) \leq L(x,y)$ for any $x \in X$, so that

$$f(x) \geq L(x,y) \geq g(y) \qquad \text{for all } x \in X,\ y \in Y.$$

Each value $g(y)$ for $y \in Y$ thus provides a lower bound to the values of the function f on X, so $\bar{\beta}$ likewise provides a lower bound, the best that can be deduced from the various values $g(y)$. This can't be more than the greatest lower bound for f on X, which by definition is $\bar{\alpha}$.

THEOREM 15 (basic characterization of saddle points). *In any two-person, zero-sum game, the following conditions on a pair $(\bar{x}, \bar{y})$ are equivalent to each other and ensure that $L(\bar{x}, \bar{y})$ is the optimal value in both player's problems.*

(a) *$(\bar{x}, \bar{y})$ is a saddle point of $L(x, y)$ on $X \times Y$.*

(b) *$\bar{x}$ is an optimal solution to problem $(\mathcal{P}_1)$, $\bar{y}$ is an optimal solution to problem $(\mathcal{P}_2)$, and the optimal values in these two problems agree.*

(c) *$\bar{x} \in X$, $\bar{y} \in Y$, and $f(\bar{x}) = g(\bar{y})$.*

Proof. The equivalence is obvious from the general inequality $f(x) \geq L(x, y) \geq g(y)$ just noted and the fact that the saddle point condition can, by its definition and that of f and g, be written as $f(\bar{x}) = L(\bar{x}, \bar{y}) = g(\bar{y})$. □

Comment. This reveals that the components $\bar{x}$ and $\bar{y}$ of a saddle point $(\bar{x}, \bar{y})$ have an independent character. They can be obtained by solving one optimization to get $\bar{x}$ and another to get $\bar{y}$. The set of saddle points is a *product set* within $X \times Y$.

Application of game theory to Lagrangians: Returning now to the case of an optimization problem $(\mathcal{P})$ in conventional format and its associated Lagrangian L on $X \times Y$, we ask what at first impression could be just a frivolous question. If we think of these elements L, X, and Y as specifying a certain two-person, zero-sum game, what would we get? As a matter of fact, in this game the strategy problem $(\mathcal{P}_1)$ for Player 1 turns out to be $(\mathcal{P})$! Indeed, the formula for the function that is to be minimized in $(\mathcal{P}_1)$,

$$f(x) := \sup_{y \in Y} L(x, y) = \sup_{y \in Y} \left\{ f_0(x) + \sum\nolimits_{i=1}^{m} y_i f_i(x) \right\},$$

gives none other than the essential objective function in $(\mathcal{P})$,

$$f(x) = \begin{cases} f_0(x) & \text{if } x \text{ is feasible,} \\ \infty & \text{if } x \text{ is not feasible.} \end{cases}$$

Thus, in tackling problem $(\mathcal{P})$ as our own, we are tacitly taking on the role of Player 1 in a certain game. Little did we suspect that our innocent project would necessarily involve us with an opponent, a certain Player 2!

This will take some time to digest, but one place to begin is with our knowledge, on the basis of game theory, that the strategy problem $(\mathcal{P}_2)$ this Player 2 is supposed to want to solve is the one in which the function $g(y) := \inf_{x \in X} L(x, y)$ is maximized over $y \in Y$. We adopt this as the problem "dual" to $(\mathcal{P})$ in the Lagrangian framework.

Lagrangian dual problem: The *dual problem* of optimization associated with $(\mathcal{P})$ (called the *primal problem* for contrast) is

$$(\mathcal{D})\qquad \begin{array}{l}\text{maximize } g(y) \text{ over all } y \in Y, \text{ where} \\ \qquad g(y) := \inf_{x\in X} L(x,y) = \inf_{x \in X} \Big\{ f_0(x) + \sum_{i=1}^m y_i f_i(x) \Big\}.\end{array}$$

Here g has the general status that f did in $(\mathcal{P})$ as the *essential* objective function in $(\mathcal{D})$, because $g(y)$ might be $-\infty$ for some choices of $y \in Y$. The feasible set in problem $(\mathcal{D})$ therefore isn't actually the set Y, but the set

$$D := \big\{ y \in Y \,\big|\, g(y) > -\infty \big\}.$$

In other words, a vector $y \in Y$ isn't regarded as a feasible solution to $(\mathcal{D})$ unless the objective value $g(y)$ is finite. Of course, closer description of D and g can't emerge until more information about X and the functions f_i has been supplied in a given case, so that the calculation of $g(y)$ can be carried out in supplementary detail.

Example: Lagrangian derivation of linear programming duality: For a linear programming problem $(\mathcal{P}_{\rm lin})$ as considered earlier, the Lagrangian is

$$\begin{aligned} L(x,y) &= \sum_{j=1}^n c_j x_j + \sum_{i=1}^m y_i \Big(b_i - \sum_{j=1}^n a_{ij} x_j \Big) \\ &= \sum_{i=1}^m b_i y_i + \sum_{j=1}^n x_j \Big(c_j - \sum_{i=1}^m y_i a_{ij} \Big) \ \text{ on } X \times Y = I\!R^n_+ \times I\!R^m_+. \end{aligned}$$

Using the second version of the formula to calculate the essential objective in $(\mathcal{D})$, we get for arbitrary $y \in I\!R^m_+$ that

$$\begin{aligned} g(y) = \inf_{x \in \boldsymbol{R}^n_+} L(x,y) &= \inf_{\substack{x_j \ge 0 \\ j=1,\dots,n}} \Big\{ \sum_{i=1}^m b_i y_i + \sum_{j=1}^n x_j \Big(c_j - \sum_{i=1}^m y_i a_{ij} \Big) \Big\} \\ &= \begin{cases} \sum_{i=1}^m b_i y_i & \text{when } c_j - \sum_{i=1}^m y_i a_{ij} \ge 0 \text{ for } j = 1,\dots,n, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The Lagrangian dual problem $(\mathcal{D})$, where $g(y)$ is maximized over $y \in Y$, comes out therefore as the previously identified dual problem $(\mathcal{D}_{\rm lin})$.

Convexity properties of the dual problem: Regardless of whether any more detailed expression is available for the feasible set D and objective function g in problem $(\mathcal{D})$, it's always true that D is a convex set with respect to which g is concave. This problem, therefore, falls within the realm of optimization problems of convex type.

Argument: Consider any points y_0 and y_1 in D along with any $\tau \in (0,1)$. Let $y_\tau = (1-\tau)y_0 + \tau y_1$. From the definition of D and g we have for each $x \in X$ that $-\infty < g(y_0) \le L(x,y_0)$ and $-\infty < g(y_1) \le L(x,y_1)$, hence

$$-\infty \;<\; (1-\tau)g(y_0) + \tau g(y_1) \;\le\; (1-\tau)L(x,y_0) + \tau L(x,y_1) = L(x,y_\tau),$$

where the equation is valid because $L(x,y)$ is affine with respect to y. Since this holds for arbitrary $x \in X$, while $g(y_\tau) = \inf_{x\in X} L(x,y_\tau)$, we obtain the concavity inequality $(1-\tau)g(y_0) + \tau g(y_1) \le g(y_\tau)$ along with the guarantee that $y_\tau \in D$.

Basic relationship between the primal and dual problems: A number of facts about $(\mathcal{P})$ and $(\mathcal{D})$ follow at once from game theory, without any need for additional assumptions. It's always true that

$$\big[\text{optimal value in } (\mathcal{P})\,\big] \;\;\ge\;\; \big[\text{optimal value in } (\mathcal{D})\,\big].$$

When these optimal values coincide, as they must by Theorem 15 if a saddle point exists for the Lagrangian L on $X \times Y$, the saddle points are the pairs $(\bar{x},\bar{y})$ such that $\bar{x}$ solves $(\mathcal{P})$ while $\bar{y}$ solves $(\mathcal{D})$. The sticking point, however is whether a saddle point does exist. For that we need to draw once more on convexity.

THEOREM 16 (duality in convex programming). *Consider an optimization problem $(\mathcal{P})$ in conventional format along with its Lagrangian dual $(\mathcal{D})$ in the case where the set X is closed and convex, the functions f_i are $\mathcal{C}^1$ and convex for $i = 0,1,\ldots,s$, and affine for $i = s+1,\ldots,m$. Then the Lagrangian $L(x,y)$ is convex in x for each $y \in Y$ as well as affine in y for each $x \in X$, and the following properties are equivalent:*

(a) *$\bar{x}$ is an optimal solution to $(\mathcal{P})$, and $\bar{y}$ is an associated Lagrange multiplier vector in the Kuhn-Tucker conditions for $(\mathcal{P})$ at $\bar{x}$;*

(b) *$(\bar{x},\bar{y})$ is a saddle point of the Lagrangian $L(x,y)$ over $X \times Y$;*

(c) *$\bar{x}$ and $\bar{y}$ are optimal solutions to $(\mathcal{P})$ and $(\mathcal{D})$, respectively, and*

$$\big[\text{optimal value in } \big(\mathcal{P}_{\text{lin}}\big)\,\big] \;\;=\;\; \big[\text{optimal value in } \big(\mathcal{D}_{\text{lin}}\big)\,\big].$$

In particular, therefore, this equation must hold if $(\mathcal{P})$ has an optimal solution for which the Kuhn-Tucker conditions are fulfilled (as is true in particular when $(\mathcal{P})$ has an optimal solution satisfying the basic constraint qualification or the refined constraint qualification that takes linear constraints into account).

Proof. Since $L(x,y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x)$, we always have $L(x,y)$ affine in y for fixed x. Because $Y = \mathbb{R}^s_+ \times \mathbb{R}^{m-s}$, the vectors $y \in Y$ have components y_i that are nonnegative for $i \in [1,s]$, and the convex programming assumptions on $(\mathcal{P})$ therefore ensure that $L(x,y)$ is convex in x when $y \in Y$. Saddle points $(\bar{x},\bar{y})$ of L over the convex product set $X \times Y$ are characterized therefore by the gradient relations $-\nabla_x L(\bar{x},\bar{y}) \in N_X(\bar{x})$ and $\nabla_y L(\bar{x},\bar{y}) \in N_Y(\bar{y})$ as noted in our discussion of saddle points. But these relations give the Lagrangian form of the Kuhn-Tucker conditions for $(\mathcal{P})$. We conclude from this that (a) and (b) are equivalent. The equivalence of (b) and (c) is based on the general principles of Theorem 15, which is applicable because $(\mathcal{P})$ and $(\mathcal{D})$ have been identified as the strategy problems for the two players in the game specified by the Lagrangian triplet L, X, Y. The final assertion of the theorem merely recalls the necessity of the Kuhn-Tucker conditions in the presence of the constraint qualifications in Theorems 11 or 12. □

Comparison with duality in linear programming: Convex programming covers linear programming as a special case, but the results in Theorem 16 are generally not as sharp or symmetric as those in Theorem 14. Nonetheless they do offer something new even for $(\mathcal{P}_{\rm lin})$ and $(\mathcal{D}_{\rm lin})$. We now know that the pairs $(\bar{x},\bar{y})$ constituting optimal solutions to these problems are precisely the saddle points over $X \times Y = \mathbb{R}^n_+ \times \mathbb{R}^m_+$ of the function

$$L(x,y) = \sum\nolimits_{j=1}^{n} c_j x_j + \sum\nolimits_{i=1}^{m} b_i y_i - \sum\nolimits_{i,j=1}^{m,n} y_i a_{ij} x_j .$$

Game-theoretic interpretation of duality: A puzzle remains. If the consideration of a problem $(\mathcal{P})$ leads inevitably to involvement in a game in which an opponent is trying to solve another problem $(\mathcal{D})$, just what is the game, and what does it signify? In principle, the game is of course as follows. Player 1, whom we identify with ourselves, chooses a vector $x \in X$ (ignoring all other constraints!) while the sinister Player 2 chooses a multiplier vector $y = (y_1,\ldots,y_m) \in Y$. Then Player 1 must pay the amount $L(x,y) = f_0(x)+y_1 f_1(x)+\ldots+y_m f_m(x)$ to Player 2. What can this mean? An important clue can usually be found in analyzing the units in which this payment is made and trying through that to interpret the quantity $L(x,y)$.

An economic example: Let's interpret $f_0(x)$ as the cost in dollars of selecting an action x from X. For $i = 1,\ldots,m$, let's take $f_i(x)$ to be the amount of resource i needed to implement x, this being measured in appropriate units (tons, hours, ...) but with the scale oriented so that $f_i(x)$ is the *excess* that would be required, relative to the amount of resource i directly available. The constraint $f_i(x) \leq 0$ refers then to staying within the available amounts. Take $(\mathcal{P})$ to be the problem of minimizing $f_0(x)$ subject to $x \in X$ and $f_i(x) \leq 0$ for $i = 1, \ldots, m$.

For units of measurement in the Lagrangian expression $L(x, y) = f_0(x) + y_1 f_1(x) + \ldots + y_m f_m(x)$ to come out consistently, the values of $L(x, y)$, like those of $f_0(x)$ must be in dollars. We deduce then that the units for the coefficient y_i must be dollars per unit of resource i. In other words, the multipliers y_i assign monetary value to the resources and act like prices!

In the game framework, Player 1 doesn't worry directly about whether the constraints $f_i(x) \leq 0$ are satisfied or not; any $x \in X$ can be selected as a "strategy." The interpretation is that resource i can be bought and sold freely on the market at a certain price y_i. If the decision x is such that $f_i(x) > 0$, so that more of resource i is needed that is already on hand, the shortfall is made up by purchasing the extra amount on the market and adding the corresponding cost $y_i f_i(x)$ to the direct cost $f_0(x)$ associated with x. On the other hand, if $f_i(x) < 0$, this means that an unused quantity $|f_i(x)|$ of resource i remains, which can be sold off to counterbalance the costs of the decision x. The proceeds $y_i |f_i(x)|$ are then subtracted from $f_0(x)$, which because of signs amounts once again to adding $y_i f_i(x)$ to $f_0(x)$. In all cases, therefore, the payment $L(x, y)$ made by Player 1 comes out as the *net cost resulting from decision x after market transactions in the resources have been taken into account, as long as the price vector is y.*

The market as the opponent: Player 2 is revealed now as the market itself, acting to keep Player 1 from making undue profit out of prices that don't reflect the true values of the resources. The market acts to set the prices so to get the highest net amount out of Player 1 that is consistent with the situation. The objective value $g(y)$ in the dual problem represents a floor to the amount the market is sure to receive.

Equilibrium prices: Our theory tells us that under mild assumptions including *convexity* there will exist a special price vector $\bar{y}$ with the property that in minimizing $L(x, \bar{y})$ over $x \in X$, as a "free market" version of the given problem $(\mathcal{P})$, there will be a solution $\bar{x}$ that actually turns out to be an optimal solution to $(\mathcal{P})$ itself. In other words, the prices $\bar{y}_i$ will have the magical property of inducing us to respect the constraints $f_i(x) \leq 0$ even though we are not obliged to! A vector $\bar{y}$ is imbued with this power if and only if it solves problem $(\mathcal{D})$.

Lagrangian relaxation: The saddle point characterization of optimal solutions to a convex programming problem $(\mathcal{P})$ in Theorem 16 has useful consequences quite generally. The "market" example just explored is only a beginning. Consider for each multiplier vector $y \in Y$ the problem

$$\text{minimize } L(x, y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) \text{ with respect to } x \in X.$$

This is called a *Lagrangian relaxation* of $(\mathcal{P})$, because it "relaxes" the constraints of type $f_i(x) \leq 0$ or $f_i(x) = 0$, incorporating them instead, to some degree, in a modified objective function; the minimization is carried out without regard to these constraints, but only to the requirement that $x \in X$. (Note that the optimal value in this relaxed problem is by definition $g(y)$.) In the situation described in Theorem 16, there is a vector $\bar{y}$ with the remarkable property that every optimal solution $\bar{x}$ to $(\mathcal{P})$ satisfies $L(x, \bar{y}) \geq L(\bar{x}, \bar{y})$ for all $x \in X$, or in other words, *the optimal set for $(\mathcal{P})$ lies within the optimal set for the relaxed problem corresponding to $\bar{y}$.* In particular, if $L(x, \bar{y})$ attains its minimum over X at just a single point $\bar{x}$ (as must be true for instance if this function is strictly convex on X, which would be the case in convex programming when f_0 is strictly convex on X), this point $\bar{x}$ has to be the unique optimal solution to $(\mathcal{P})$!

If only there were some way of determining such a multiplier vector $\bar{y}$, we could in principle avoid the pain of minimizing $f_0(x)$ over the feasible set C in $(\mathcal{P})$ and merely minimize $L(x, \bar{y})$ over the presumably much simpler set X. In fact the determination of $\bar{y}$ is not so far-fetched. Theorem 16 tells us that in looking for $\bar{y}$ we are looking for an optimal solution to the dual problem $(\mathcal{D})$.

Practical realization: Lagrangian relaxation is typically built into procedures involving a whole sequence of multiplier vectors $y^\nu \in Y$. In each iteration, $L(x, y^\nu)$ is minimized over $x \in X$ to get a point x^ν in the hope that the sequence in X so generated, or perhaps an auxiliary sequence constructed out of it, will be asymptotically optimal in $(\mathcal{P})$. The vectors y^ν aren't specified in advance, but produced from information revealed as the computations go on. The scheme is designed with the aim of getting the y^ν's to form an optimizing sequence for $(\mathcal{D})$.

Troubles in the absence of convexity: It mustn't be overlooked that the success of such an approach depends on $(\mathcal{P})$ being a *convex* programming problem. When that isn't true, there's no reason to believe that an optimal solution $\bar{y}$ to problem $(\mathcal{D})$, or for that matter any other vector $\bar{y}$, never mind how cleverly chosen, will have the property required. Without that, any confidence in a method of Lagrangian relaxation being able to "solve" problem $(\mathcal{P})$ is definitely misplaced.

Lower bounds on the primal optimal value: For any dual feasible solution $y \in D$, the finite value $g(y)$, obtained by solving the relaxed problem of minimizing $L(x, y)$ over $x \in X$, is a lower bound to the optimal value in $(\mathcal{P})$. This follows from general relationship between the optimal values in $(\mathcal{P})$ and $(\mathcal{D})$ that was noted just before Theorem 16. In many situations, regardless of convexity, the ability to generate such a lower bound is helpful in gaining information about how near x might be to solving $(\mathcal{P})$. Such information might enter into a stopping criterion in a numerical method for solving $(\mathcal{P})$.

Decomposition of large-scale problems: When problems of optimization are very large in dimension—with thousands of variables and possibly a great many constraints to go with them—there's strong interest in setting up methods of computation might take advantage of special structure that might be present. An attractive idea is that of breaking a problem $(\mathcal{P})$ down into much smaller subproblems, to be solved independently, maybe by processing them in parallel.

Of course, there's no way of achieving such a decomposition once and for all, because that would presuppose knowing unrealistically much about the problem before it's even been tackled. Rather, one has to envision schemes in which an initial decomposition of $(\mathcal{P})$ into subproblems yields information leading to a better decomposition into modified subproblems, and so forth iteratively in a manner that can be justified as eventually generating an asymptotically optimal sequence $\{x^\nu\}_{\nu \in \mathbb{N}}$. As long as $(\mathcal{P})$ is well posed, such a sequence can be deemed to solve it in the sense of Theorem 2.

Example: Decentralization of decision making. Consider now a situation where a number of separate agents or decision makers, indexed by $k = 1, \dots, r$, would solve problems of optimization independent of each other if it weren't for the necessity of sharing certain resources. Agent k would prefer just to minimize $f_{0k}(x_k)$ over all x_k in a set $X_k \subset \mathbb{R}^{n_k}$, but there are mutual constraints

$$f_{i1}(x_1) + \cdots + f_{ir}(x_r) \le c_i \;\text{ for }\; i = 1, \dots, m$$

requiring some coordination. Furthermore, there's community interest in having the coordination take place in such a way that the overall sum $f_{01}(x_1) + \cdots + f_{0r}(x_r)$ is kept low. In other words, the problem as seen from the community is to

$$(\mathcal{P}^*) \qquad \begin{aligned} &\text{minimize }\; f_0(x) := f_{01}(x_1) + \cdots + f_{0r}(x_r) \;\text{ subject to} \\ &\qquad f_i(x) := f_{i1}(x_1) + \cdots + f_{ir}(x_r) - c_i \le 0 \;\text{ for }\; i = 1, \dots, m, \\ &\qquad x := (x_1, \dots, x_r) \in X := X_1 \times \cdots \times X_r. \end{aligned}$$

To concentrate attention on the nicest case (much could still be said under weaker assumptions), suppose that the sets X_k are convex, closed and bounded, and the functions f_{ik} are convex and of class $\mathcal{C}^1$, with f_{0k} strictly convex. Then $(\mathcal{P}^*)$ is a well posed, convex programming problem in which the objective function f_0 is strictly convex. Provided that $(\mathcal{P}^*)$ has a feasible solution, which we henceforth assume as well, $(\mathcal{P}^*)$ has a unique optimal solution $\bar{x}$; this follows from Theorems 1 and 8(b). Assuming further that this optimal solution $\bar{x}$ satisfies the basic constraint qualification in Theorem 11 or the refined one in Theorem 12, we know from Theorem 16 that an optimal solution $\bar{y}$ exists for the associated dual problem $(\mathcal{D}^*)$. In terms of $\bar{y}$, as discussed under the heading of Lagrangian relaxation, $\bar{x}$ is displayed as the unique optimal solution to the problem of minimizing $L(x, \bar{y})$ over $x \in X$.

What is the form of this relaxed problem? For any $y \in Y$ we have

$$\begin{aligned} L(x, y) : &= \sum\nolimits_{k=1}^{r} f_{0k}(x_k) + \sum\nolimits_{i=1}^{m} y_i \Big(\sum\nolimits_{k=1}^{r} f_{ik}(x_k) - c_i \Big) \\ &= \sum\nolimits_{k=1}^{r} L_k(x_k, y) - y{\cdot}c \\ \text{with } & L_k(x_k, y) := f_{0k}(x_k) + \sum\nolimits_{i=1}^{m} y_i f_{ik}(x_k), \; c = (c_1, \ldots, c_m). \end{aligned}$$

Thus, to minimize $L(x, \bar{y})$ over $x \in X$ is the same as to

$$\text{minimize } L_1(x_1, \bar{y}) + \cdot + L_r(x_r, \bar{y}) \text{ over all } (x_1, \ldots, x_r) \in X_1 \times \cdots \times X_r.$$

But in this subproblem there is no interaction between the different components x_k: the minimum is achieved by solving the separate problems

$$(\mathcal{P}_k(\bar{y})) \qquad \text{minimize } L_k(x_k, \bar{y}) \text{ over all } x_k \in X_k \qquad (k = 1, \ldots, r).$$

Implication: The vector $\bar{y}$ has the amazing property of allowing the decomposition of the community problem $(\mathcal{P}^*)$ into a number of subproblems, which can be dealt with in parallel by the separate agents. For $k = 1, \ldots, r$, Agent k solves $(\mathcal{P}_k(\bar{y}))$ and thereby determines an element $\bar{x}_k$. These elements need only then be put together as $(\bar{x}_1, \ldots, \bar{x}_r)$ to get the unique optimal solution $\bar{x}$ to $(\mathcal{P})$. Thus, the optimal set of decisions $\bar{x}_k$ for the community as a whole can be achieved without either a meeting to coordinate the conflicting aims of the agents or a "budget tsar" to parcel out the resource amounts c_i. Instead, the agents can come up with these decisions by solving individual problems of optimization from their individual points of view. All that is needed is a certain way of incorporating the resource circumstances into the individual objectives through terms $\bar{y}_i f_{ik}(x_k)$.

Economic interpretation: This fits perfectly with the market interpretation of Lagrange multipliers that was given earlier. The coefficient $\bar{y}_i$ is the market price in dollars per unit of resource i. In contemplating the selection of $x_k \in X_k$, Agent k faces the direct cost $f_{0k}(x_k)$. But Agent k knows also that if x_k requires an amount $f_{ik}(x_k) > 0$ of resource i, such amount can be obtained at cost $\bar{y}_i f_{ik}(x_k)$, whereas if x_k produces an amount of resource i, i.e., if $f_{ik}(x_k) < 0$, such amount can be sold for a return of $\bar{y}_i[-f_{ik}(x_k)]$ against the costs. From this "free market" perspective, Agent k is occupied simply with minimizing $f_{0k}(x_k) + \sum_{i=1}^m \bar{y}_i f_{ik}(x_k)$ over $x_k \in X_k$, which is problem $(\mathcal{P}_k(\bar{y}))$. The idea that the individual agents, in pursuing merely their own interests in this way, will end up making the decisions optimal for the community as a whole, is of course classical economics in the vein of Adam Smith. It's no wonder, then, that mathematical economists have intensely studied this kind of result in optimization theory.

Remark: This example was worth exploring at length because it illustrates not only an important tactic, which can be used to counter the high dimensionality of large-scale problems when solving them, but the importance of optimization theory in mathematical modeling itself. Optimization is more than just numerical optimization. It serves in analyzing and understanding various situations as well as in calculating answers to problems with specific data.

Abuses of the notion of price decomposition: Free-market zealots have captured the world's attention with the idea that economic behavior that's optimal from society's interest can be induced simply by letting markets set the proper prices. For instance, if every form of pollution has its price, and the price is right, companies will keep pollution within society's desired bounds just by acting out of their own self interest. But the underpinnings of this assertion rest on the structure of the economy being "convex." Roughly speaking, that's only true in the classical setting of an economy consisting only of a huge mass of infinitesimal agents, not the modern setting of several major agents in each industry. Abuses of this notion of decomposition can arise also, politics aside, when people who simply want a convenient way of solving a problem $(\mathcal{P})$ don't appreciate the crucial assumptions.

Some other approaches to problem decomposition:

Frank-Wolfe decomposition: For problems of minimizing a nonlinear, differentiable convex function f_0 over a set C specified by linear constraints, a possible way to generate a sequence of feasible solutions x^ν from an initial point $x^0 \in C$ as follows. Having arrived at x^ν, form the linearized function

$$l^\nu(x) := f_0(x^\nu) + \big\langle \nabla f_0(x^\nu),\, x - x^\nu \big\rangle$$

and minimize l^ν over C to get a point $\hat{x}^\nu$. The information provided by this subproblem can be used in various ways. For instance, $d^\nu = \hat{x}^\nu - x^\nu$ turns out to be a descent vector for f_0, moreover one giving a feasible direction into C at x^ν (unless x^ν itself already minimizes l^ν over C, in which case x^ν already has to be optimal). We won't go into the details of the possibilities here. The main thing is that the subproblem of minimizing l^ν over C is one of linear programming, in contrast to that of minimizing f_0 over C. In special situations, such as when C is a box or a even just a product of polyhedral sets of low dimension, it breaks down into still smaller subproblems solvable in closed form or in parallel.

Benders decomposition: This term was originally attached to a scheme in linear programming, but the concept can be explained much more generally. Imagine in $(\mathcal{P})$ that the vector x is split into two vector components, $x = (x', x'')$ with $x' \in I\!R^{n'}$ and $x'' \in I\!R^{n''}$, representing the "hard" and "easy" parts of $(\mathcal{P})$ in the following sense. For any *fixed* choice of x', the residual problem of minimizing $f_0(x', x'')$ over all x'' such that $(x', x'') \in C$ is "easy" because of its special structure. Let $\varphi(x')$ denote the optimal value in this subproblem, with x' as parameter, and let B denote the set of all x' for which at least one x'' exists with $(x', x'') \in C$. In principle then, $(\mathcal{P})$ can be solved by minimizing $\varphi(x')$ over all $x' \in B$ to get $\bar{x}'$, and then solving the "easy" subproblem associated with this $\bar{x}'$ to get $\bar{x}''$; the pair $\bar{x} = (\bar{x}', \bar{x}'')$ will be optimal in $(\mathcal{P})$. Once more, this is just the skeleton of an approach which has to be elaborated into an iterative procedure to make it practical. Typically, duality enters in representing φ approximately in such a manner that it can minimized effectively.

Decomposition through block coordinate minimization: Sometimes $(\mathcal{P})$ has many "easy" parts if they could be treated separately. Suppose $x = (x_1, \ldots, x_r)$ with vector components $x_k \in I\!R^{n_k}$; each x_k designates a *block* of coordinates of x in general. For any $\hat{x} \in C$, denote by $(\mathcal{P}_k(\hat{x}))$ for $k = 1, \ldots, r$ the subproblem

$$\text{minimize } f_0(\hat{x}_1, \ldots, \hat{x}_{k-1}, x_k, \hat{x}_{k+1}, \ldots, x_r) \text{ over all } x_k \text{ such that}$$
$$(\hat{x}_1, \ldots, \hat{x}_{k-1}, x_k, \hat{x}_{k+1}, \ldots, x_r) \in C.$$

The basic idea is to generate a sequence $\{x^\nu\}_{\nu=0}^\infty$ in C from an initial x^0 as follows. Having reached x^ν, choose an index k^ν in $\{1, \ldots, r\}$ and solve subproblem $(\mathcal{P}_{k^\nu}(x^\nu))$, obtaining a vector $\bar{x}_{k^\nu}$. Replace the k^ν-component of x^ν by this vector, leaving all the other components as they were, to obtain the next point $x^{\nu+1}$. Obviously, if this is to work well, care must be exercised that the same index in $\{1, \ldots, r\}$ isn't always chosen; some scheme in which every index repeatedly gets its turn is essential. But what's not realized by many people is that, although the method may lead to lower values of f_0, but never higher, it can get hung up and fail to produce an optimal sequence x^ν unless strong assumptions are fulfilled. Not only must f_0 be differentiable, it must be strictly convex, and the feasible set C be convex and have the product form $C_1 \times \cdots \times C_r$.