

October 1993

LIDS-P-2204

Appears in J.O.T.A., Vol. 89, 1996, pp. 1-15.

THEVENIN DECOMPOSITION¹ AND LARGE-SCALE OPTIMIZATION

by

Dimitri P. Bertsekas²

Abstract

Thevenin's theorem, one of the most celebrated results of electric circuit theory, provides a two-parameter characterization of the behavior of an arbitrarily large circuit, as seen from two of its terminals. We interpret the theorem as a sensitivity result in an associated minimum energy/network flow problem, and we abstract its main idea to develop a decomposition method for convex quadratic programming problems with linear equality constraints, such as those arising in a variety of contexts such as Newton's method, interior point methods, and least squares estimation. Like Thevenin's theorem, our method is particularly useful in problems involving a system consisting of several subsystems, connected to each other with a small number of coupling variables.

Keywords: optimization, decomposition, circuit theory, network flows.

¹ This research was supported by NSF under Grant CCR-91-03804.

² Professor, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139.

1. INTRODUCTION

This paper is motivated by a classical result of electric circuit theory, Thevenin's theorem,³ that often provides computational and conceptual simplification of the solution of electric circuit problems involving linear resistive elements. The theorem shows that, when viewed from two given terminals, such a circuit can be described by a single arc involving just two electrical elements, a voltage source and a resistance (see Fig. 1). These elements can be viewed as sensitivity parameters, characterizing how the current across the given terminals varies as a function of the external load to the terminals. They can be determined by solving two versions of the circuit problem, one with the terminals open-circuited and the other with the terminals short-circuited (by solution of a circuit problem, we mean finding the currents and/or the voltages across each arc). Mathematically, one can interpret Thevenin's theorem as the result of systematic elimination of the circuit voltages and currents in the linear equations expressing Kirchhoff's laws and Ohm's law. Based on this interpretation, one can develop multidimensional versions of Thevenin's theorem (Refs. 5-7).

In this paper we interpret the ideas that are implicit in Thevenin's theorem within an optimization context, and we use this interpretation to develop a decomposition method for quadratic programs with linear constraints. Significantly, these are the types of problems that arise in the context of interior point methods for linear programming, and more generally in the context of constrained versions of Newton's method. Our method is not entirely novel, since it is based on the well-known partitioning (or Benders decomposition) approach of large-scale optimization. However, the partitioning idea is applied here in a

³ Leon Thevenin (1857-1926) was a French telegraph engineer. He formulated his theorem at the age of 26. His discovery met initially with skepticism and controversy within the engineering establishment of the time. Eventually the theorem was published in 1883. A brief biography of Thevenin together with an account of the development of his theorem is given by C. Suchet in Ref. 1. For a formal statement of the theorem and a discussion of its applications in circuit theory, see for example Refs. 2-4.

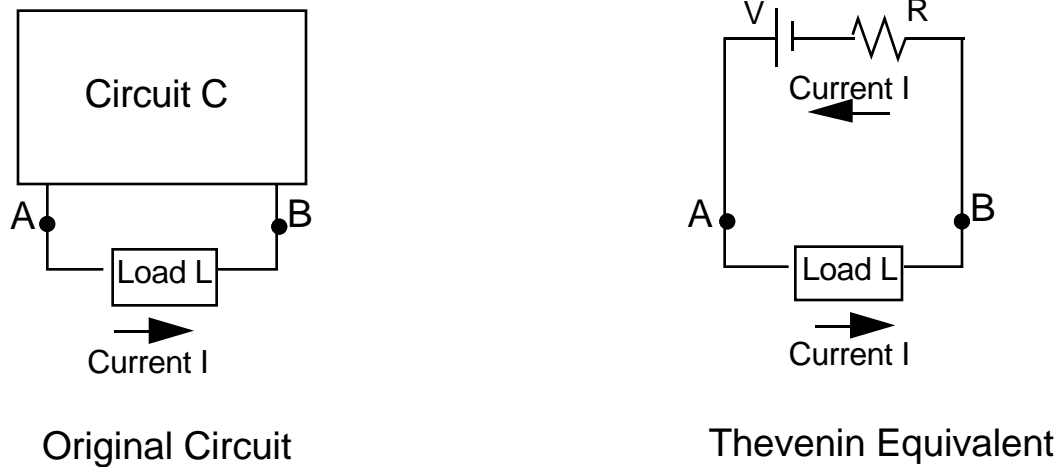


Figure 1: Illustration of Thevenin's theorem. A linear resistive circuit C acts on a load connected to two of its terminals A and B like a series connection of a voltage source V and a resistance R . The parameters V and R depend only on the circuit C and not on the load, so if in particular the load is a resistance L , the current drawn by the load is

$$I = \frac{V}{L + R}.$$

The parameters V and R are obtained by solving the circuit for two different values of L , for example $L = \infty$, corresponding to open-circuited terminals, and $L = 0$ corresponding to short-circuited terminals.

way that does not seem to have been exploited in the past.

Our interpretation is based on a well-known relation between the solution of linear resistive electric circuit problems and minimum energy optimization problems (Refs. 3,8). In particular, consider a linear resistive electric network with node set \mathcal{N} and arc set \mathcal{A} . Let v_i be the voltage of node i and let x_{ij} be the current of arc (i, j) . Kirchhoff's current law says that for each node i , the total outgoing current is equal to the total incoming current

$$\sum_{\{j|(i,j) \in \mathcal{A}\}} x_{ij} = \sum_{\{j|(j,i) \in \mathcal{A}\}} x_{ji}. \quad (1)$$

Ohm's law says that the current x_{ij} and the voltage drop $v_i - v_j$ along each arc (i, j) are related by

$$v_i - v_j = R_{ij}x_{ij} - t_{ij}, \quad (2)$$

2. The General Decomposition Framework

where $R_{ij} \geq 0$ is a resistance parameter and t_{ij} is another parameter that is nonzero when there is a voltage source along the arc (i, j) (t_{ij} is positive if the voltage source pushes current in the direction from i to j).

Consider also the problem

$$\min \sum_{(i,j) \in \mathcal{A}} \left(\frac{1}{2} R_{ij} x_{ij}^2 - t_{ij} x_{ij} \right) \quad (3a)$$

$$\text{s.t.} \quad \sum_{\{j|(i,j) \in \mathcal{A}\}} x_{ij} = \sum_{\{j|(j,i) \in \mathcal{A}\}} x_{ji}, \quad \forall i \in \mathcal{N}. \quad (3b)$$

(The quadratic cost above has an energy interpretation.) Then it can be shown that a set of currents $\{x_{ij} \mid (i, j) \in \mathcal{A}\}$ and voltages $\{v_i \mid i \in \mathcal{N}\}$ satisfy Kirchhoff's current law and Ohm's law, if and only if $\{x_{ij} \mid (i, j) \in \mathcal{A}\}$ solve problem (3) and $\{v_i \mid i \in \mathcal{N}\}$ are Lagrange multipliers corresponding to the Kirchhoff's current law constraints (1). The proof consists of showing that Kirchhoff's current law and Ohm's law constitute necessary and sufficient optimality conditions for problem (3).

In view of the relation just described, it is clear that Thevenin's theorem can alternatively be viewed as a sensitivity result for a special type of quadratic programming problem. In Section 2 we use an elimination or partitioning approach to develop this result for general convex quadratic programming problems with linear equality constraints. In Sections 3 and 4 we delineate circumstances for which our methodology is most likely to be fruitfully applied. In particular, in Section 3 we consider network flow problems consisting of loosely connected subnetworks, while in Section 4 we consider separable problems with nearly block-diagonal constraint matrix and coupling variables, such as those arising in a number of large-scale problem contexts, including stochastic programming.

2. THE GENERAL DECOMPOSITION FRAMEWORK

Our starting point is the problem

$$\min F(x) + G(y) \quad (4a)$$

$$\text{s.t.} \quad Ax + By = c, \quad x \in X, y \in Y. \quad (4b)$$

2. The General Decomposition Framework

Here $F : \mathfrak{R}^n \rightarrow \mathfrak{R}$ and $G : \mathfrak{R}^m \rightarrow \mathfrak{R}$ are convex functions, X and Y are convex subsets of \mathfrak{R}^n and \mathfrak{R}^m , respectively, and A is an $r \times n$ matrix, B is an $r \times m$ matrix, and $c \in \mathfrak{R}^r$ is a given vector. The optimization variables are x and y , and they are linked through the constraint $Ax + By = c$.

We have primarily in mind problems with special structure, where substantial simplification or decomposition would result if the variables x were fixed. Accordingly, we consider eliminating y by expressing its optimal value as a function of x . This approach is well-known in the large-scale mathematical programming literature, where it is sometimes referred to as *partitioning* or *Benders decomposition*. In particular, we first consider optimization with respect to y for a fixed value of x , that is,

$$\min G(y) \tag{5a}$$

$$\text{s.t. } By = c - Ax, \quad y \in Y, \tag{5b}$$

and then minimize with respect to x . Suppose that an optimal solution, denoted $y(Ax)$, of this problem exists for each $x \in X$. Then if x^* is an optimal solution of the problem

$$\min F(x) + G(y(Ax)) \tag{6a}$$

$$\text{s.t. } x \in X, \tag{6b}$$

it is clear that $(x^*, y(Ax^*))$ is an optimal solution of the original problem (4). We call problem (6) the *master problem*.

Let us assume that problem (5) has an optimal solution and at least one Lagrange multiplier for each $x \in X$, that is, a vector $\lambda(Ax)$ such that

$$\min_{y \in Y, By=c-Ax} G(y) = \max_{\lambda \in \mathfrak{R}^r} q(\lambda, Ax) = q(\lambda(Ax), Ax), \tag{7}$$

where $q(\cdot, Ax)$ is the corresponding dual functional given by

$$q(\lambda, Ax) = \inf_{y \in Y} \{G(y) + \lambda'(Ax + By - c)\} = \tilde{q}(\lambda) + \lambda'Ax, \tag{8}$$

with

$$\tilde{q}(\lambda) = \inf_{y \in Y} \{G(y) + \lambda'(By - c)\}. \tag{9}$$

2. The General Decomposition Framework

Then the master problem (6) can also be written as

$$\min F(x) + Q(Ax) \tag{10a}$$

$$\text{s.t. } x \in X, \tag{10b}$$

where

$$Q(Ax) = \max_{\lambda \in \mathbb{R}^r} q(\lambda, Ax). \tag{11}$$

It is possible to characterize the differentiability properties of Q in terms of Lagrange multipliers of problem (5). In particular, using Eq. (7), one can show that if the function \tilde{q} of Eq. (9) is strictly concave over the set $\{\lambda \mid \tilde{q}(\lambda) > -\infty\}$, then Q is differentiable at Ax , and $\nabla Q(Ax)$ is equal to the unique Lagrange multiplier $\lambda(Ax)$ of problem (5) [which is also the unique maximizer of $q(\lambda, Ax)$ in Eq. (7)]. This result can form the basis for an iterative gradient-based solution of the master problem (6).

In this paper we propose an alternative approach, which is based on calculating the function $Q(Ax)$ in closed form. A prominent case where this is possible is when the minimization problem above is quadratic with equality constraints, as shown in the following proposition.

Proposition 2.1: Assume that the matrix B has rank r , and that

$$G(y) = \frac{1}{2}y'Ry + w'y, \quad Y = \{y \mid Dy = d\},$$

where R is a positive definite $m \times m$ matrix, D is a given matrix, and d, w are given vectors. Assume further that the constraint set $\{y \mid By = c - Ax, Dy = d\}$ is nonempty for all x . Then the function Q of Eq. (11) is given by

$$Q(Ax) = \frac{1}{2}(Ax - b)'M(Ax - b) + \gamma, \tag{12}$$

where M is a $r \times r$ positive definite matrix, γ is a constant, and

$$b = c - B\bar{y}, \tag{13}$$

with

$$\bar{y} = \arg \min_{y \in Y} G(y). \tag{14}$$

Furthermore, the vector

$$\lambda(Ax) = M(Ax - b), \quad (15)$$

is the unique Lagrange multiplier of the problem

$$\min G(y) \quad (16a)$$

$$\text{s.t. } By = c - Ax, \quad y \in Y. \quad (16b)$$

Proof: We first note that because R is positive definite and the constraint set $\{y \mid By = c - Ax, Dy = d\}$ is nonempty for all x , problem (5) has a unique optimal solution and at least one Lagrange multiplier vector. We have by definition [cf. Eq. (8)]

$$\begin{aligned} q(\lambda, Ax) &= \min_{Dy=d} \left\{ \frac{1}{2}y'Ry + w'y + \lambda'(Ax + By - c) \right\} \\ &= \min_{Dy=d} \left\{ \frac{1}{2}y'Ry + (w + B'\lambda)'y \right\} + \lambda'(Ax - c). \end{aligned}$$

Let us assume without loss of generality that D has full rank; if it doesn't, we can replace the constraint $Dy = d$ by an equivalent full rank constraint and the following analysis still goes through.

By a well-known quadratic programming duality formula, we have

$$\begin{aligned} q(\lambda, Ax) &= \max_{\mu} \left\{ -\frac{1}{2}\mu'DR^{-1}D'\mu - (d + DR^{-1}(w + B'\lambda))'\mu \right\} \\ &\quad - \frac{1}{2}(w + B'\lambda)'R^{-1}(w + B'\lambda) + \lambda'(Ax - c). \end{aligned}$$

The maximum above is attained at

$$\mu(\lambda) = -(DR^{-1}D')^{-1}(d + DR^{-1}(w + B'\lambda))$$

and by substitution in the preceding equation, we obtain

$$\begin{aligned} q(\lambda, Ax) &= -\frac{1}{2}(d + DR^{-1}(w + B'\lambda))'(DR^{-1}D')^{-1}(d + DR^{-1}(w + B'\lambda)) \\ &\quad - \frac{1}{2}(w + B'\lambda)'R^{-1}(w + B'\lambda) + \lambda'(Ax - c). \end{aligned} \quad (17)$$

Thus we can write

$$q(\lambda, Ax) = -\frac{1}{2}\lambda'M^{-1}\lambda - \lambda'b + \lambda'Ax + \gamma \quad (18)$$

2. The General Decomposition Framework

for an appropriate positive definite matrix M , vector b , and constant γ . The unique Lagrange multiplier $\lambda(Ax)$ maximizes $q(\lambda, Ax)$ over λ , so from Eq. (18) we obtain

$$\lambda(Ax) = M(Ax - b) \tag{19}$$

and by substitution in Eq. (18),

$$Q(Ax) = q(\lambda(Ax), Ax) = \frac{1}{2}(Ax - b)'M(Ax - b) + \gamma. \tag{20}$$

There remains to show that b is given by Eqs. (13) and (14). From Eq. (18), b is the gradient with respect to λ of $q(\lambda, Ax)$, evaluated at $\lambda = 0$ when $x = 0$, that is,

$$b = -\nabla_{\lambda}q(0, 0). \tag{21}$$

Since $q(\lambda, 0)$ is the dual functional of the problem

$$\min G(y) \tag{22a}$$

$$\text{s.t. } Dy = d, By = c, \tag{22b}$$

Eqs. (13) and (14) follow from a well-known formula for the gradient of a dual function.

Q.E.D.

Note that the preceding proposition goes through with minor modifications under the weaker assumption that R is positive definite over the nullspace of the matrix D , since then G is strictly convex over Y .

One approach suggested by Proposition 2.1 is to calculate M and b explicitly [perhaps using the formulas (17) and (18) of the proof of Proposition 2.1], and then to solve the master problem (6) for the optimal solution x^* using Eqs. (12)-(15). This is the method of choice when the inversions in Eqs. (17) and (18) are not prohibitively complicated. However, for many problems, these inversions are very complex; an important example is when the problem

$$\begin{aligned} \min \quad & \frac{1}{2}y'Ry + w'y \\ \text{s.t.} \quad & By = c - Ax, Dy = d \end{aligned}$$

3. Application to Network Optimization

involves a network as in the examples discussed in the next section. In such cases it may be much preferable to solve the problems (14) and (16) by an iterative method, which, however, cannot produce as a byproduct M and b via the formulas (17) and (18).

An alternative approach, which apparently has not been suggested earlier, is to solve the problem

$$\min \frac{1}{2}y'Ry + w'y \quad (23a)$$

$$\text{s.t. } Dy = d \quad (23b)$$

to obtain the vector b [cf. Eqs. (13) and (14)], and then solve certain quadratic programs to obtain the matrix M . In particular, suppose that the matrix A has rank r , and suppose that we solve r problems of the form (16) with x equal to each of r vectors such that the corresponding vectors $Ax - b$ are linearly independent. Then based on the relation $\lambda(Ax) = M(Ax - b)$ [cf. Eq. (15)], the Lagrange multipliers of these problems together with b yield the matrix M . This approach is particularly attractive for problems where the dimension of x is relatively small, and subproblems of the form (14) and (16) are best solved using an iterative method. An added advantage of an iterative method in the present context is that the final solution of one problem of the form (16) typically provides a good starting point for solution of the others. In the sequel we will restrict ourselves to this second approach.

3. APPLICATION TO NETWORK OPTIMIZATION

Let us apply the decomposition method just described to network optimization problems with convex separable quadratic cost problem

$$\min \sum_{(i,j) \in \mathcal{A}} \left(\frac{1}{2}R_{ij}x_{ij}^2 - t_{ij}x_{ij} \right) \quad (24a)$$

$$\text{s.t. } \sum_{\{j|(i,j) \in \mathcal{A}\}} x_{ij} = \sum_{\{j|(j,i) \in \mathcal{A}\}} x_{ji}, \quad \forall i \in \mathcal{N}, \quad (24b)$$

3. Application to Network Optimization

where R_{ij} is a given positive scalar and t_{ij} is a given scalar. Such problems arise in an important context. In particular, the quadratic programming subproblems of interior point methods, as applied to *linear* network optimization problems with bound constraints on the arc flows, are of this type. The same is true for the subproblems arising when barrier methods or multiplier methods are used to eliminate the bound constraints of differentiable convex network flow problems and Newton's method is used to solve the corresponding "penalized" subproblems.

Let us first show that Thevenin's theorem can be derived as the special case of our decomposition approach where x consists of the current of a single arc.

Example 3.1: Derivation of Thevenin's Theorem.

Let us fix a specific arc (\bar{i}, \bar{j}) of the network, and let us represent by x the arc flow $x_{\bar{i}\bar{j}}$, and by y the vector of the flows of all arcs other than (\bar{i}, \bar{j}) , that is,

$$y = \{x_{ij} \mid (i, j) \neq (\bar{i}, \bar{j})\}.$$

Then the coupling constraint is $Ax + By = c$, where $c = 0$, $A = 1$, and B is a row vector of zeroes, ones, and minus ones, where the ones correspond to the outgoing arcs from node \bar{i} , except for arc (\bar{i}, \bar{j}) , and the minus ones correspond to the incoming arcs to node \bar{i} . Calculating explicitly the function $G(Ax)$ using the formula (17) is complicated, so we follow the approach of computing b and M . To apply this approach, we should solve two problems:

- (1) The corresponding problem (23). This is the same as the original network optimization problem (24) but with the conservation of flow constraints corresponding to nodes \bar{i} and \bar{j} eliminated.
- (2) The corresponding problem (5) with the flow $x_{\bar{i}\bar{j}}$ fixed at zero; this amounts to removing arc (\bar{i}, \bar{j}) .

These two problems will give us M and b , which are scalars because x is one-dimensional in this example. The corresponding master problem is

$$\min \frac{1}{2}R_{\bar{i}\bar{j}}x_{\bar{i}\bar{j}}^2 - t_{\bar{i}\bar{j}}x_{\bar{i}\bar{j}} + \frac{1}{2}Mx_{\bar{i}\bar{j}}^2 - Mbx_{\bar{i}\bar{j}} \tag{25a}$$

$$\text{s.t. } x_{\bar{i}\bar{j}} \in \mathfrak{R} \tag{25b}$$

3. Application to Network Optimization

so the optimal value of $x_{\bar{i}\bar{j}}$ is

$$x_{\bar{i}\bar{j}}^* = \frac{t_{\bar{i}\bar{j}} + Mb}{R_{\bar{i}\bar{j}} + M}. \quad (26)$$

The above expression is precisely the classical Thevenin's theorem. To see this, we recall the connection between the quadratic program (24) and linear resistive electric circuit problems given in the introduction. Then we can view both the original problem (25) as well as the corresponding subproblems (1) and (2) above as electric circuit problems. These subproblems are:

- (1) The original circuit problem with arc (\bar{i}, \bar{j}) short-circuited. By Eq. (13), b is the short-circuit current of arc (\bar{i}, \bar{j}) .
- (2) The original circuit problem with arc (\bar{i}, \bar{j}) removed or open-circuited. By Eq. (15), Mb is the open-circuit voltage drop $v_{\bar{i}} - v_{\bar{j}}$ across arc (\bar{i}, \bar{j}) .

Consider the circuit C obtained from the original after arc (\bar{i}, \bar{j}) is removed. Equation (26) shows that, when viewed from the two terminals \bar{i} and \bar{j} , circuit C can be described by a single arc involving just two electrical elements, a voltage source Mb and a resistance M . This is Thevenin's theorem.

Example 3.2: Decomposition Involving Two Subnetworks.

Note that in the preceding example, the parameters b and M depend only on the characteristics of the subnetwork C and the nodes \bar{i} and \bar{j} , and not on the characteristics of the arc (\bar{i}, \bar{j}) . In particular, given two subnetworks C_1 and C_2 , connected at just two nodes A and B (see Fig. 2), one of the subnetworks, say C_1 , can be replaced by its equivalent two-parameter arc, and the resulting subnetwork can be solved to determine the flows within C_2 , as well as the flow going from C_1 to C_2 , which in turn can be used to finally obtain the flows within C_1 . The problem involving the interconnection of C_1 and C_2 , can be solved by solving smaller-dimensional problems as follows (see Fig. 2):

- (a) Two problems involving just C_1 to determine its two-parameter Thevenin representation.
- (b) One problem involving just C_2 to find the flow x^* going from C_1 to C_2 , as well as the flows within C_2 .

3. Application to Network Optimization

- (c) One problem involving just C_1 and the flow x^* to determine the flows within C_1 .

Note that the computational effort to solve a quadratic network problem is generically proportional to the cube of its size, so if C_1 and C_2 are of comparable size, the overall computational effort is cut in half through the use of Thevenin's theorem. Furthermore, the two parameters describing C_1 can be reused if C_2 is replaced by a different subnetwork.

Alternatively we can represent as vector x the flow going from C_1 to C_2 , and as y the set of flows of the arcs of C_1 and C_2 , and apply our general decomposition method. Then, to determine x^* through the corresponding parameters M and b of Eq. (12)-(15), it is necessary to solve two problems, one with the terminals shortcircuited and another with $x = 0$ (see Fig. 3). Each of these problems involves the solution of two independent subproblems, one involving subnetwork C_1 and the other involving subnetwork C_2 . However, one additional problem involving C_1 and another involving C_2 must now be solved to determine the flows within C_1 and C_2 using the value of x^* . Still the computational effort is smaller than the one required to solve the original network problem without decomposition.

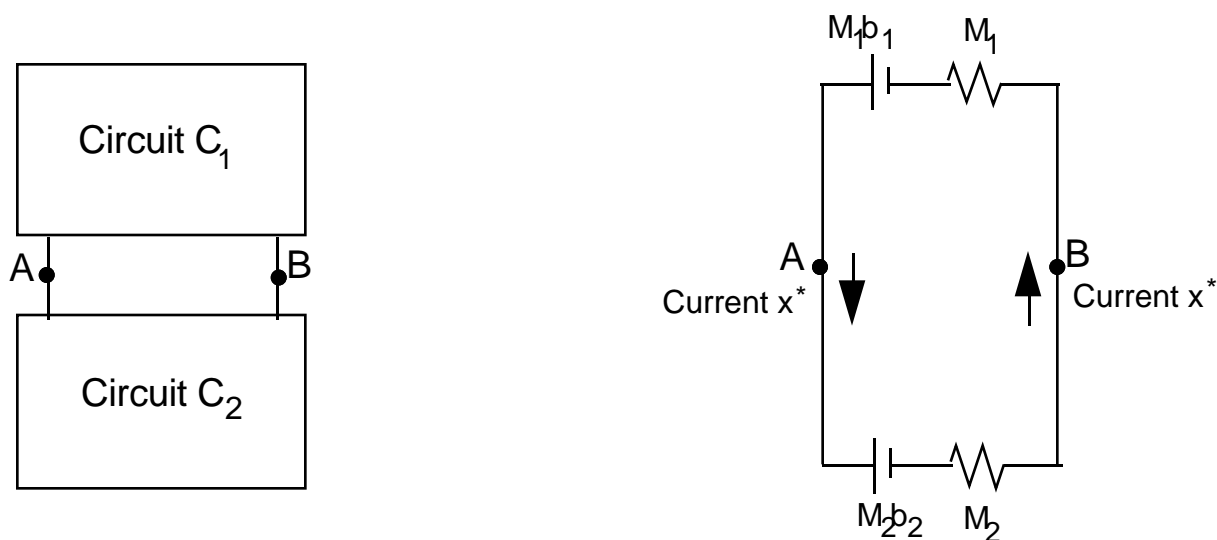


Figure 3: The optimal flow going from subnetwork C_1 to subnetwork C_2 can be found by constructing the two-parameter representations of C_1 and C_2 , and solving for the current x^* in the equivalent two-arc network.

3. Application to Network Optimization

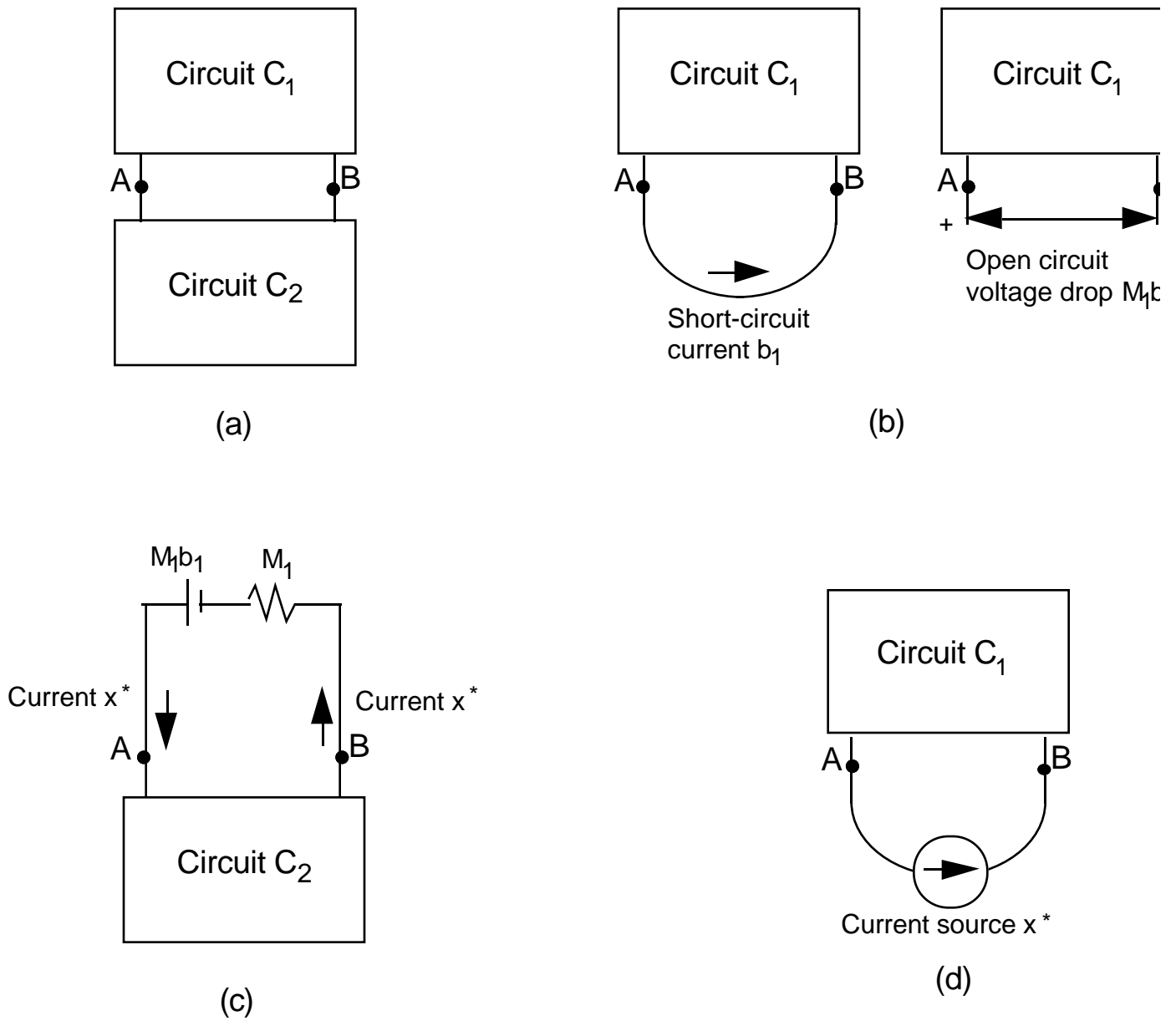


Figure 2: Using Thevenin's theorem for decomposition of a problem involving the solution of two subnetworks connected at two nodes as in (a). The Thevenin parameters M_1 and b_1 are first obtained by solving two subnetwork problems involving C_1 as in (b). Then the problem involving C_2 and the Thevenin equivalent of C_1 is solved to obtain the flow x^* going from C_1 to C_2 as in (c). This flow is used to find the currents in C_1 as in (d).

Example 3.3: Decomposition Involving Several Subnetworks.

The decomposition procedure outlined in Figs. 2 and 3 can be extended to the case of several

3. Application to Network Optimization

subnetworks. Consider the quadratic network optimization problem (24), and suppose that we select n arcs (i_k, j_k) , $k = 1, \dots, n$, such that after these arcs are deleted, the corresponding quadratic network optimization problem simplifies substantially. Then we can consider $(n + 1)$ quadratic network optimization problems: one corresponding to problem (24) with all arcs (i_k, j_k) , $k = 1, \dots, n$, shortcircuited, and n others corresponding to the flow of one arc (i_k, j_k) fixed at a nonzero value, and the other $(n - 1)$ arcs deleted. The first of these problems will yield the n dimensional vector b of Eq. (13), and the remaining problems will yield the matrix M using Eq. (15). The optimal flows of the arcs (i_k, j_k) , $k = 1, \dots, n$, are then obtained by solving the corresponding n -dimensional master program (6).

This procedure can be useful in problems involving networks that consist of component networks that are interconnected through a few arcs. The subproblems obtained by deleting these arcs separate into simpler component problems, and once the parameters of the components are determined, the final solution is obtained by solving a quadratic program involving just the flows of the interconnecting arcs.

There are potential computational savings with this approach, which can be quantified if we assume a specific interconnection structure. For example, assume that the network consists of $(n + 1)$ subnetworks such that each pair of subnetworks (C_k, C_{k+1}) , $k = 1, \dots, n$, is connected by $(m + 1)$ arcs. Assume that the number of nodes of all subnetworks is bounded by a constant N . Then the computation involved if we ignore the structure of the problem is $O(n^3(N + m)^3)$. Suppose that we apply the decomposition method outlined in the present example. Then, $(n + 1)$ problems of size N must be solved to find the vector b , and mN problems involving two adjacent subnetworks and two interconnecting arcs, as per Example 2, must also be solved to obtain the matrix M for a total of $O(mnN^3)$ computation. For large N , the Thevenin decomposition approach is substantially superior if m is substantially less than n^2 .

The approach illustrated in the preceding example may also be fruitfully applied to more complex network problems, involving for example arcs with gains and multiple commodities, as long as the basic structure of loosely connected subnetworks is present.

4. SEPARABLE QUADRATIC PROGRAMS WITH COUPLING VARIABLES

Consider the quadratic program

$$\min \frac{1}{2}x'Px - v'x + \sum_{j=1}^J \left(\frac{1}{2}y_j'R_jy_j + w_j'y_j \right) \quad (27a)$$

$$\text{s.t. } Ex = e, \quad A_jx + B_jy_j = c_j, \quad D_jy_j = d_j, \quad j = 1, \dots, J, \quad (27b)$$

where $x \in \Re^n$, the matrices and vectors above have appropriate dimensions, and P and R_j are positive definite. This block-diagonal constraint structure with coupling variables arises frequently in practice. For example, it arises in stochastic programming problems, that is, two-stage stochastic optimal control problems, for which there is much recent interest. In the context of these latter problems, x is the decision variable of the first stage, and y_j is the decision variable of the second stage, corresponding to the j th (stochastic) outcome of the first stage. Our methodology is applicable to linear stochastic programming problems, when they are approached via interior point methods, since the subproblem to be solved in each iteration of the interior point method is a quadratic program of the form (27).

By representing (y_1, \dots, y_J) by the vector y , we can apply our decomposition method. The matrix M in this example is block-diagonal with blocks denoted M_j , $j = 1, \dots, J$, and this results in great simplification. The method consists of the following steps:

- (a) Find the optimal solution \bar{y}_j of each of the J subproblems

$$\begin{aligned} \min \quad & \frac{1}{2}y_j'R_jy_j + w_j'y_j \\ \text{s.t.} \quad & D_jy_j = d_j, \end{aligned}$$

to obtain the vector $b = (b_1, \dots, b_J)$ from

$$b_j = c_j - B_j\bar{y}_j, \quad j = 1, \dots, J,$$

[cf. Eq. (13)].

4. Separable Quadratic Programs with Coupling Variables

- (b) Find the columns of each block M_j , $j = 1, \dots, J$, by solving subproblems of the form

$$\begin{aligned} \min \quad & \frac{1}{2}y_j'R_jy_j + w_j'y_j \\ \text{s.t.} \quad & A_jx + B_jy_j = c_j, \quad D_jy_j = d_j \end{aligned}$$

for a sufficient number of values of A_jx , and by using the corresponding Lagrange multipliers $\lambda_j(A_jx)$ in the relation

$$\lambda_j(A_jx) = M_j(A_jx - b_j).$$

It is necessary to solve r_j such subproblems, where r_j is the dimension of M_j .

- (c) Find x by solving the master program (6), which has the form

$$\begin{aligned} \min \quad & \frac{1}{2}x'Px - v'x + \frac{1}{2} \sum_{j=1}^J (A_jx - b_j)'M_j(A_jx - b_j) \\ \text{s.t.} \quad & Ex = e. \end{aligned}$$

When the constraint $Ex = e$ is absent ($E = 0$ and $e = 0$), the optimal solution of the master problem is given in closed form as

$$x^* = \left(P + \sum_{j=1}^J A_j' M_j A_j \right)^{-1} \sum_{j=1}^J (v + A_j' M_j b_j). \quad (28)$$

When x is one-dimensional, the optimal solution x^* can also be represented as the solution of a resistive electric circuit as shown in Fig. 4.

We note that the idea of partitioning of linear-quadratic problems through the use of coupling variables has been applied recently by Chang et. al. (Ref. 9) and Ralph (Ref. 10) in the context of discrete-time linear-quadratic optimal control, and by Nikoukhah et. al. (Ref. 11) in the context of least-squares estimation. In these methods, the time horizon is partitioned in blocks, with coupling between the blocks provided by the state vectors at the block boundaries (these boundary state vectors constitute the vector x in our context). The subproblems needed to obtain the matrices M_j and vectors B_j corresponding to the blocks are linear-quadratic optimal control or estimation problems with fixed initial and/or terminal state, which can be solved in parallel.

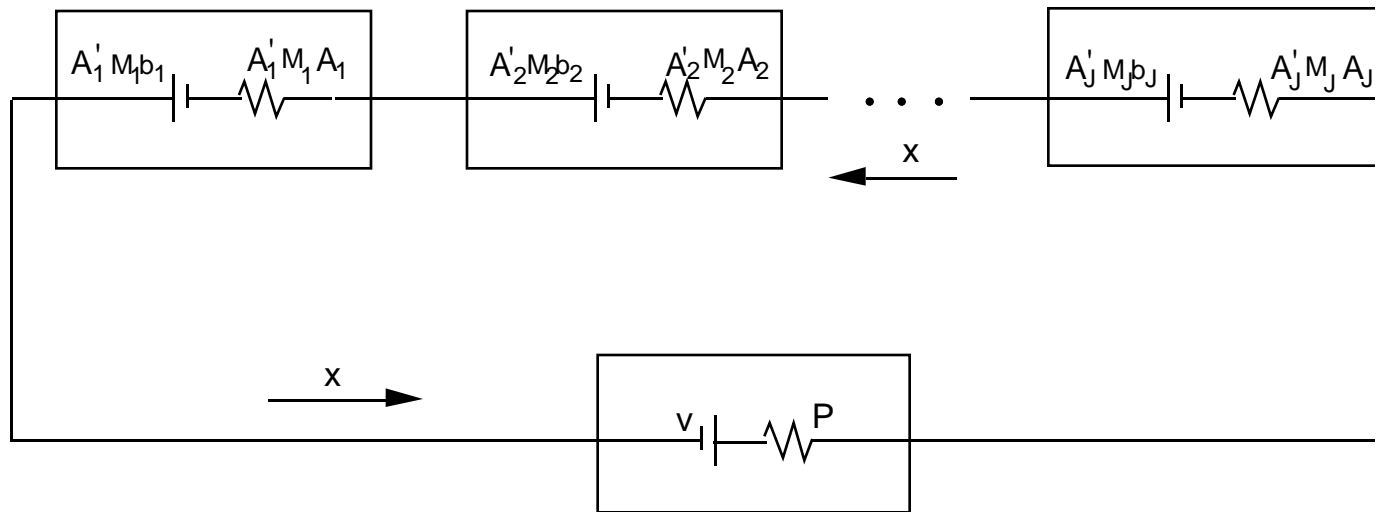


Figure 4: Electrical circuit analog of the optimal solution x^* of the master problem of Eq. (28) for the case where x is one-dimensional.

REFERENCES

1. Suchet, C., Electrical Engineering, Vol. 68, pp. 843-844, 1949.
2. Desoer, C. A., and Kuh, E. S., Basic Circuit Theory, McGraw Gill, New York, New York, 1969.
3. Guillemin, E. A., Theory of Linear Physical Systems, J. Wiley, New York, New York, 1963.
4. Van Valkenburg, M. E., and Kinariwala, B. K., Linear Circuits, Prentice-Hall, Englewood Cliffs, New Jersey, 1982.
5. Corazza, G. C., Someda, C. G., and Longo, G., Generalized Thevenin's Theorem for Linear n -Port Networks, IEEE Transactions on Circuit Theory, Vol. 16, pp. 564-566, 1969.
6. Director, S. W., and Wayne, D. A., Computational Efficiency in the Determination of the Thevenin and Norton Equivalents, IEEE Trans. on Circuit Theory, Vol. 19, pp. 96-98, 1972.
7. Hajj, I. N., Computation of Thevenin and Norton Equivalents, Electronic Letters, Vol.

12, 1970.

8. Dennis, J. B., *Mathematical Programming and Electrical Circuits*, Technology Press of M.I.T., Cambridge, Massachusetts, 1959.

9. Chang, S. C., Chang, T. S., and Luh, P. B., A Hierarchical Decomposition for Large-Scale Optimal Control Problems with Parallel Processing Structure, *Automatica*, Vol. 25, pp. 77-86, 1989.

10. Ralph, D., A Parallel Method for Discrete-Time Optimal Control Problems, Technical Report CTC93TR118, Advanced Computing Research Institute, Cornell University, Feb. 1993.

11. Nikoukhah, R., Taylor, D., Levy, B. C., and Willsky, A. S., Graph Structure and Recursive Estimation of Noisy Linear Relations, Rapport de Recherche No. 1912, INRIA, 1993.