TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2013

 | ISBN 978-0-9843378-4-2
http://dx.doi.org/10.1287/educ.2013.0116

Sparse Solutions to Complex Models

Xin Chen
Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, xinchen@illinois.edu

Chung Piaw Teo
Department of Decision Sciences, National University of Singapore, Singapore 119077, bizteocp@nus.edu.sg

Abstract Recent years witnessed the proliferation of the notion of sparsity and its applications in operations research models. To bring to the attention and raise the interest of the operations research community on this topic, we present in this tutorial a wide range of complex models that admit sparse yet effective solutions. Our examples range from compressed sensing and process flexibility to queuing applications, and from equation systems and optimization problems to game theory models.

Keywords sparse; complex; equations; quadratic program; game theory; two-stage stochastic program; flexibility

1. Introduction

In operations research, we often need to make delicate trade-offs of various factors, leading to extremely complex models. How to make sense out of these complex models is practically important while intellectually challenging. Interestingly and many times surprisingly, simple yet effective solutions exist for some of these complex models. By "simple," we refer to solutions that are characterized by only a few key factors among hundreds or even thousands of factors, i.e., solutions that are "sparse."

The existence of effective sparse solutions is important in several aspects. First, it significantly reduces the computational complexity of searching for effective solutions. Several examples we present are NP-hard in the worst case. However, with the help of sparsity, they can be solved by polynomial time algorithms under some conditions either in the worst case or in some average sense. Second, it allows one to identify key factors that affect system performance and thus develop simple and effective strategies to enhance the performance of a system.

In this tutorial, we provide several examples to illustrate a wide range of such models. Our first example focuses on identifying a sparest solution of an undetermined linear system of equations. Though the general problem is NP-hard, one can show that such a solution can be found by solving a linear program under some conditions. This result has been proven important for compressed sensing theory and applications. In our second example, we analyze random standard quadratic programs (StQPs) that are widely used for resource allocation models and portfolio selection models. It is shown under some probability models that any global optimal solution is sparse. When the entries of the Hessian of the objective function are independently generated by a distribution that is concave in its support, the probability that any optimal solution has more than k nonzero elements decays exponentially in k. When the Hessian is a Gaussian orthogonal ensemble (GOE), with a very high probability, any optimal solution has no more than two nonzero elements when the order of the Hessian is large. Similar observations carry over to two-player matrix games. Specifically,

under some probability models, in a Nash equilibrium in mixed strategies, each player uses very few—in fact, with a high probability no more than two—pure strategies.

Effective sparse solutions also exist for some two-stage stochastic programs, which find important applications in process flexibility and risk-sharing networks. In this setting, sparse first-stage solution may not be globally optimal but can be shown to be provably near optimal in many natural environments. In the area of manufacturing network design, it is by now well known that a little flexibility introduced into the manufacturing process—properly configured and orchestrated across the network—can significantly improve the network's ability to match manufacturing capacity with volatile product demands. The automobile industry, for instance, has moved away from *focused* plants (where one plant produces essentially one product) to partially flexible plants (where one plant produces a small number of products). The advantage of flexible plant is illustrated by the following:

> "The initial investment is slightly higher, but long-term costs are lower in multiples," said Chris Bolen, manager of Ford's Windsor engine plant, which uses the flexible system to machine new three-valve-per-cylinder heads for Ford's 5.4-liter V8 engine . . . . Ford says the system will help it meet changes in demand. "If our business was hit by a significant down sizing from V8s to V6s or V6s to (four-cylinder engines) or diesels in North America, we'll be able to react to that without years of turnaround," said Kevin Bennett, Ford director of power train manufacturing. "It's essential we be able to react to the market more rapidly than in the past." (Phelan [45])

Indeed, we have seen a proliferation of the notion of sparsity and its applications in operations research problems in recent years. Instead of providing a comprehensive survey, the purpose of this tutorial is to bring to the attention and raise the interest of operations research community on the possibility of simple solutions to a host of complex models.

The organization of this chapter is as follows. In §2, we present sparsity results for undetermined linear systems of equations. In §3, we illustrate that random standard quadratic programs admit sparse solutions, followed by sparsity results on Nash equilibria in mixed strategies in two-player random matrix games in §4. In §5, we show that a class of two-stage stochastic programs admits approximate sparse solutions. Finally, we conclude this chapter with additional applications of sparse solutions in other related areas in §6.

2. Undetermined Linear Systems

Consider an undetermined linear system of equations

$$Ax = b,$$

where A is an $m \times n$ real matrix with $m < n$ and $b \in \Re^m$. Assume that A has a full row rank. The linear system admits infinite number of solutions. An interesting question is to find one that is the sparsest among all solutions; that is, we would like to solve the following optimization problem:

$$\begin{aligned} \min \quad & \|x\|_0 \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{1}$$

Here, for $x \in \Re^n$, $\|x\|_0$ denotes the number of nonzero elements of x. Later we will also use $\|x\|_p$ to denote the p-norm of vector x for $p > 0$; i.e., $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$.

Problem (1) has attracted a great deal of attention in recent years in the area of compressed sensing (Eldar and Kutyniok [30]). In that literature, $x \in \Re^n$ is the raw data representing the object of interest such as signals and images, A is a sensing matrix with each row representing a linear measurement of the raw data, and $b \in \Re^m$ is the outcome of the linear measurements. Essentially, through the linear measurements, the high-dimensional data are compressed into a much lower dimensional space, which allows for significantly reduced cost and improved efficiency of computation and storage. A critical question is under what conditions the original data can be reconstructed accurately and efficiently.

Since we face an undetermined linear system of equations, there is no way to pin down to a unique solution without imposing additional conditions on the solution. Fortunately, in practice, raw data are usually sparse, i.e., they have sparse representations under some appropriately chosen basis, or compressible, i.e., they can be well approximated by sparse representations. Thus, one can impose sparsity conditions on desirable solutions. In particular, problem (1) attempts to find a sparsest solution given a sensing matrix and the outcome of the linear measurements.

Problem (1) is a challenging combinatorial optimization problem. Indeed, it is NP-hard in general (Donoho [29]). To deal with this challenge, Chen et al. [18] propose to solve a closely related problem, referred to as the l_1 minimization problem:

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{2}$$

Clearly, the above problem can be easily reformulated as a linear programming problem and thus can be solved in a polynomial time. Surprisingly, Donoho [29] and Candès and Tao [17] prove that the optimal solutions for problem (2) and problem (1) are identical in many instances when the original data are sufficiently sparse. In the following, we introduce one such result. Our presentation follows Candès [16] and Baraniuk et al. [3].

Definition 1 (Candès and Tao [17]). A matrix $A \in \Re^{m\times n}$ is said to have the *restricted isometry property* (RIP) *with order* k, if there exists a constant $\delta_k \in (0,1)$ such that for any vector $x \in \sum_k$,

$$(1-\delta_k)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta_k)\|x\|_2^2, \tag{3}$$

where $\sum_k = \{x \in \Re^n \colon \|x\|_0 \le k\}$.

If a full row rank matrix $A \in \Re^{m\times n}$ satisfies the RIP with order $2k$, then problem (1) admits at most one feasible solution with no more than k nonzero elements. Indeed, if both x and x' are feasible and belong to $\sum_k$, then $z = x - x'$ belongs to $\sum_{2k}$ and satisfies $Az = 0$. From the RIP of A with order $2k$, $z = x - x' = 0$. Thus, a signal with no more than k nonzero elements can be exactly recovered.

Theorem 1 (Candès [16]). *Assume that* $\delta_{2k} < \sqrt{2} - 1$. *Given any feasible solution of problem* (1), x, *the solution* x^* *to problem* (2) *obeys*

$$\|x^* - x\|_1 \le C_0 \|x - x_k\|_1$$

and

$$\|x^* - x\|_2 \le C_0 k^{-1/2} \|x - x_k\|_1$$

for some constant C_0, *where* $x_k \in \sum_k$ *is a vector derived from* x *by setting* x*'s smallest* $(n-k)$ *elements to be zero.*

Clearly, if the original data x have at most k nonzero elements, the above theorem implies that solving problem (2) leads to an optimal solution for problem (1).

To construct matrices with good RIPs, one often has to refer to randomization. Specifically, we construct a matrix $A \in \Re^{m\times n}$ by generating its elements independently from an identical probability distribution such that

$$P\big(\|Ax\|_2^2 \ge (1+\epsilon)\|x\|_2^2\big) \le 2e^{-mc_0(\epsilon)}, \tag{4}$$

where the probability is taken over the random matrix A, and $c_0(\epsilon)$ is a positive constant only depending on $\epsilon \in (0,1)$. One notable distribution that satisfies the condition (4) is Gaussian $\mathcal{N}(0, 1/n)$.

Theorem 2 (Baraniuk et al. [3]). *Suppose that* n, m, k, *and* $\delta_k \in (0,1)$ *are given. If the elements of matrix* $A \in \Re^{m\times n}$ *are generated independently from an identical probability*

distribution satisfying condition (4)*, then there exist some positive constants* c_1 *and* c_2 *depending only on* δ_k *such that* A *satisfies* (3) *for any* $k \le c_1 m/\ln(n/k)$ *with probability* $1-2e^{-c_2 m}$*.*

Theorems 1 and 2 imply that when the elements of matrix A are independently identically distributed according to Gaussian $\mathcal{N}(0, 1/n)$, with overwhelming probability (i.e., with probability exponentially decaying in m), the optimal solutions for problem (2) and problem (1) are identical.

3. Standard Quadratic Programs

In this section, we focus on the StQP in which one minimizes a quadratic objective function subject to a simplex constraint as follows:

$$\begin{aligned} \min \quad & x^T Q x \\ \text{s.t.} \quad & e^T x = 1, \quad x \ge 0, \end{aligned} \tag{5}$$

where $Q = [Q_{ij}] \in \mathcal{S}^n$, $e \in \Re^n$ is the all 1-vector and $\mathcal{S}^n$ denotes the set of symmetric real matrices with order n.

This problem appears in numerous applications such as resource allocation (Irabaki and Katoh [37]) and portfolio selection (Markowitz [42]). It covers also other problems such as the maximal clique problem in discrete optimization (Gibbons et al. [33]) and determining the copositivity of a matrix in linear algebra (Bomze et al. [11]). Since the StQPs are among the simplest quadratic programs and are useful to model a variety of applications, they have attracted the attention of researchers in various fields, and a host of algorithms have been proposed in the literature. For details, we refer to recent papers (Bomze et al. [11], Scozzari and Tardella [49], Yang and Li [56]) and the references therein.

Since determining the copositivity of a matrix is NP-hard and can be reduced to solving a StQP, we know StQP is NP-hard in the worst case. It is intriguing to understand whether solving StQPs would be any easier in some average sense. In addition to the tractability concern, our interest on problem (5) was also partially motivated by the exciting advance on the undetermined linear systems of equations with sparse solutions. To see the linkage, consider the problem

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b, \quad x \ge 0, \end{aligned} \tag{6}$$

which is the same as problem (2) with additional nonnegativity constraints on the decision variables. Denote a nontrivial sparest solution of problem (6) by x^* satisfying $\rho = \|x^*\|_1 > 0$. Then x^* must be the sparest solution of the following least square optimization problem:

$$\begin{aligned} \min \quad & \|Ax - b\|_2^2 \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = \rho, \quad x \ge 0. \end{aligned} \tag{7}$$

Let $c = A^T b$. Because of the special constraints in the above problem, we have

$$b^T A x = c^T x = \frac{1}{\rho} x^T (e c^T) x = \frac{1}{2\rho} x^T (e c^T + c e^T) x.$$

Ignoring the constant term in the objective, we can rewrite problem (7) as the following:

$$\begin{aligned} \min \quad & x^T \left(A^T A - \frac{1}{\rho}(e c^T + c e^T) \right) x \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = \rho, \quad x \ge 0. \end{aligned} \tag{8}$$

Note that the above problem is homogeneous in x. Without loss of generality, we can replace the constraints by $e^T x = 1$, $x \geq 0$, and thus problem (8) reduces to a special case of the StQP model (5). The above interesting relation and the established results on l_1 minimization problems indicate that for some StQPs, global optimum solutions are expected to be very sparse.

Unlike the results in the previous section, which are based on the assumption that sparse solutions exist, we will illustrate that problem (5) admits a sparse global optimal solution when Q is randomly generated from some distributions. Specifically, we assume that Q satisfies the following conditions.

Assumption 1. *The random matrix Q satisfies the following conditions*:
(a) *Q is symmetric.*
(b) *Q's diagonal elements are independently and identically distributed with cumulative distribution function $F(\cdot)$.*
(c) *Q's strict upper triangular elements are independently and identically distributed with cumulative distribution function $G(\cdot)$ and are also independent of its diagonal elements.*

We start with a simple setting in which the elements of Q in problem (5) are generated by a discrete distribution with a positive probability at the left end point of its support.

Theorem 3. *Suppose Assumption* 1 *holds with $F = G$. In addition, G has a mass with probability $p_0 > 0$ at the (finite) left end point of its support. Let x^* be a sparsest global optimal solution of problem* (5)*. Then the probability that x^* has only one nonzero element is*

$$P(\|x^*\|_0 = 1) \geq 1 - (1 - p_0)^n.$$

The above theorem implies that as n goes to ∞, with a high probability there exists an optimal solution with one positive element, and the optimal objective value is given by the minimal diagonal element of Q. Interestingly, this result is not valid anymore if the assumption in Theorem 3 is violated, as we demonstrate in the following.

Theorem 4. *Suppose Assumption* 1 *holds with $F = G$. If G is continuous, then*

$$P(\|x^*\|_0 \geq 2) \geq \frac{n-1}{2n-1}.$$

Though the above theorem states that $P(\|x^*\|_0 \geq 2) \geq (n-1)/(2n-1) \to \frac{1}{2}$ as n goes to ∞, we will show in the following that $P(\|x^*\|_0 \geq k)$ decays exponentially in k for a class of random matrices.

We start by exploring the optimality conditions for problem (5). Note that since the matrix Q is indefinite in most cases, there might exist multiple optimal solutions to problem (5). If there exists a global optimal solution x^* to problem (5) with only one nonzero element ($\|x^*\|_0 = 1$), then such a solution can be identified easily by comparing all the diagonal elements of Q. Therefore, in what follows we concentrate on the cases where the sparest global optimal solution x^* to problem (5) has more than one nonzero elements.

Proposition 1 (Chen et al. [21]). *Suppose that x^* is one of the sparsest global optimal solutions of problem* (5) *satisfying $\|x^*\|_0 = k > 1$. Let $Q_{\mathcal{K}} \in \mathcal{S}^{k\times k}$ be the principal submatrix of Q induced by the index set of all the nonzero elements of x^*, and define $\lambda^* = (x^*)^T Q x^*$. Then the following conclusions hold*:
C.1. *There exists a row (or column) of $Q_{\mathcal{K}}$ such that the average of all its elements is strictly less than the minimal diagonal element of Q.*
C.2. *The matrix $Q_{\mathcal{K}} - \lambda^* E_k$ is positive semidefinite.*

Property C.1 in the above proposition follows from the first-order optimality condition, whereas property C.2 is derived from the second-order optimality condition. The following sparsity results of the global optimal solution are mainly built upon property C.1 and the probability bounds in Theorem 6.

Theorem 5 (Chen et al. [21]). *Suppose Assumption 1 holds. If $G(\cdot)$ is concave in its support and there exists some $\alpha \in (0,1]$ such that $G(x) \geq \alpha F(x)$ for any $x \in (-\infty, \infty)$, then for any global optimal solution x^* of problem (5),*

$$P(\|x^*\|_0 \geq k) \leq \tau^{k-1}\left(\frac{1}{(1-\tau)^2} + \frac{k-1}{1-\tau}\right) + \frac{n(n+1)}{2}\tau^{\lfloor\sqrt{2n\alpha}\rfloor}, \tag{9}$$

where $\tau = (1/(1+\alpha/2))^{1/2}$.

Theorem 5 implies that the probability $P(\|x^*\|_0 \geq k)$ is bounded above by a function that decays exponentially in k for reasonably large k. However, for small k, the right-hand side bound in relation (9) might turn out to be larger than 1. More accurate estimates can be obtained by invoking a more careful analysis. In fact, we show in Chen et al. [21] that the probability $P(\|x^*\|_0 = 2)$ is no more than $\frac{7}{12}$ under the same assumption of Theorem 5.

To shed some light on how Theorem 5 is derived, we need the following probability bound on order statistics.

Let U_r, $r = 1, \ldots, n$ be independent continuous random variables each with a cumulative distribution $F(\cdot)$ and $u_1 \leq u_2 \leq \cdots \leq u_n$ be the order statistics of U_r's. Let V_r, $r = 1, 2, \ldots, n$ be independent continuous random variables each with a cumulative distribution $G(\cdot)$ and $v_1 \leq v_2 \leq \cdots \leq v_n$ be the order statistics of V_r's. Assume that the random vectors $[U_r]_{r=1}^n$ and $[V_s]_{s=1}^n$ are independent from each other. Define

$$\rho(n,k) = P\left(\sum_{r=1}^{k} u_r \leq k v_1\right).$$

Theorem 6 (Chen et al. [21]). *If $G(\cdot)$ is concave in its support and there exists some $\alpha \in (0,1]$ such that $G(x) \geq \alpha F(x)$ for any $x \in (-\infty, \infty)$, then we have*

$$\rho(n,k) \leq \begin{cases} \left(\dfrac{1}{1+\alpha/2}\right)^{k/2} & \text{if } k \leq \lfloor\sqrt{2n\alpha}\rfloor, \\ \left(\dfrac{1}{1+\alpha/2}\right)^{\sqrt{2n\alpha}/2} & \text{otherwise.} \end{cases} \tag{10}$$

In addition,

$$\sum_{i=1}^{n} P\left(\sum_{r=1}^{k} u_r \leq (k+1)v_1 - v_i\right) \leq (k+1)\rho(n,k). \tag{11}$$

To see how the above inequalities are used to establish Theorem 5, denote $Q_{i,:}$ the ith row of the matrix Q and $\bar{Q}_{i,:}$ the sorted sequence in increasing order consisting of all the elements in Q_{ij} for $i \neq j$. Define the following probability events

$$\mathcal{H}_i^k = \left\{Q_{ii} + \sum_{j=1}^{k-1} \bar{Q}_{ij} \leq k \min_{j=1:n} Q_{jj}\right\}, \quad i = 1, \ldots, n, \quad \mathcal{H}^k = \bigcup_{i=1}^{n} \mathcal{H}_i^k. \tag{12}$$

If a sparsest global optimal solution to problem (5) has exactly k positive elements with $k > 1$, it follows from the second conclusion of Proposition 1 that there exists a row of the submatrix $Q_{\mathcal{K}}$ whose average is strictly less than $\min_{j=1:n} Q_{jj}$, where $\mathcal{K}$ is the index set of nonzero elements of x^*. This implies that there exists a row, say, i, such that

$$Q_{ii} + \sum_{i=1}^{k-1} \bar{Q}_{ij} \leq k \min_{j=1:n} Q_{jj}.$$

Therefore,

$$\begin{aligned} P(\|x^*\|_0 \geq k) &\leq P\Big(\bigcup_{i=1}^{n} \mathcal{H}_i^k\Big) \\ &\leq \sum_{i=1}^{n} P(\mathcal{H}_i^k) \\ &= \sum_{i=1}^{n} P\Big(\sum_{j=1}^{k-1} \bar{Q}_{ij} \leq k \min_{j=1:\,n} Q_{jj} - Q_{ii}\Big) \\ &= \sum_{i=1}^{n} P\Big(\sum_{j=1}^{k-1} u_j \leq k v_1 - v_i\Big) \\ &\leq k\rho(n, k-1), \end{aligned}$$

where u and v are the order statistics defined earlier. Theorem 5 now follows directly from the inequalities established in Theorem 6.

One significant restriction of the above result is that the cumulative distribution function $G(\cdot)$ is required to be concave, or, equivalently, its probability density function is nonincreasing in its support. Though it covers notable distributions such as uniform and exponential, it excludes normal distributions, widely used in the literature of l_1 minimization. In what follows we show how to employ both properties C.1 and C.2 to significantly improve the probability bounds in Theorem 5 and, more importantly, quantify the sparsity of global optimal solutions of StQPs generated from normal distributions.

To present our analysis, let $Q_{\mathcal{K}}$ be the principal submatrix induced by the index set $\mathcal{K}$. Define the following probability events:

$$\mathcal{H}_1^{\mathcal{K}} = \big\{Q_{\mathcal{K}} \text{ satisfies C.1}\big\}; \tag{13}$$

$$\mathcal{H}_{1,\,i}^{\mathcal{K}} = \Big\{\sum_{j\in\mathcal{K}} Q_{ij} \leq k \min_{j=1:n} Q_{jj}\Big\}; \tag{14}$$

$$\tilde{\mathcal{H}}_{1,\,i}^{\mathcal{K}} = \Big\{\sum_{j\in\mathcal{K}} Q_{ij} \leq k \min_{j\notin\mathcal{K}\setminus\{i\}} Q_{jj}\Big\}; \tag{15}$$

$$\mathcal{H}_2^{\mathcal{K}} = \big\{\exists\,\lambda^* \in \Re\colon Q_{\mathcal{K}} - \lambda^* E_k \succeq 0\big\}. \tag{16}$$

Let x^* be one of the sparsest global optimal solutions of problem (5). From Proposition 1, we have that if $\|x^*\|_0 = k$, then there exists an index set $\mathcal{K}$ with $|\mathcal{K}| = k$, which satisfies properties C.1 and C.2. Therefore, for any given index set $\mathcal{K}$ with $|\mathcal{K}| = k$,

$$P(\|x^*\|_0 = k) \leq C(n,k) P\big(\mathcal{H}_1^{\mathcal{K}} \cap \mathcal{H}_2^{\mathcal{K}}\big),$$

where $C(n,k) = n!/((n-k)!\,k!)$ denotes the binomial coefficient. Because $\mathcal{H}_1^{\mathcal{K}} = \bigcup_{i\in\mathcal{K}} \mathcal{H}_{1,\,i}^{\mathcal{K}}$ and any principal submatrix of a positive semidefinite matrix is still positive semidefinite, we have that

$$\mathcal{H}_1^{\mathcal{K}} \cap \mathcal{H}_2^{\mathcal{K}} \subseteq \bigcup_{i\in\mathcal{K}} \big(\mathcal{H}_{1,\,i}^{\mathcal{K}} \cap \mathcal{H}_2^{\mathcal{K}\setminus\{i\}}\big).$$

Since $\mathcal{H}_{1,\,i}^{\mathcal{K}} \subseteq \tilde{\mathcal{H}}_{1,\,i}^{\mathcal{K}}$ and the two events $\tilde{\mathcal{H}}_{1,\,i}^{\mathcal{K}}$ and $\mathcal{H}_2^{\mathcal{K}\setminus\{i\}}$ are independent, we have, for any given $i \in \mathcal{K}$,

$$P(\|x^*\|_0 = k) \leq kC(n,k) P(\tilde{\mathcal{H}}_{1,\,i}^{\mathcal{K}}) P(\mathcal{H}_2^{\mathcal{K}\setminus\{i\}}). \tag{17}$$

Thus, it remains to bound the probabilities $P(\tilde{\mathcal{H}}_{1,\,i}^{\mathcal{K}})$ and $P(\mathcal{H}_2^{\mathcal{K}\setminus\{i\}})$.

Under Assumption 1 with $F = G$, if G is continuous and concave in its support, then an upper bound of the probability $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$ can be readily derived using Theorem 6. We can also show that

$$P(\exists\, \lambda \in \Re\colon Q - \lambda E \succeq 0) \leq \frac{2^n}{(n+1)!},$$

which allows us to establish an upper bound for $P(\mathcal{H}_2^{\mathcal{K}\setminus\{i\}})$. By combining those inequalities, we have the following result.

Theorem 7 (Chen and Peng [20]). *Suppose Assumption 1 holds with $F = G$ and $G(\cdot)$ is continuous and concave in its support. Let x^* be one global optimal solution of problem (5). It holds for $k \geq 2$ that*

$$P(\|x^*\|_0 = k) \leq \frac{(k-1)^{k-1}2^{k-1}}{((k-1)!)^2}. \tag{18}$$

We remark that compared with the bound in Theorem 5, for small k, the bound in the above theorem is not stronger. However, for reasonably large k, the new bound is tighter. Indeed, the Stirling's approximation implies that the new bound is roughly in the order of $\exp(-O(k \ln k))$, whereas the one in Theorem 5 is $\exp(-O(k))$.

All of the results established so far require the concavity of the cumulative distribution function G. We now focus on the case in which Q is a GOE; i.e., G and F are the cumulative distribution functions of $\mathcal{N}(0, 1/2)$ and $\mathcal{N}(0, 1)$, respectively.

Again, our task is to bound the probabilities $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$ and $P(\mathcal{H}_2^{\mathcal{K}\setminus\{i\}})$. To bound the probability $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$, we need the following key technical result, whose proof is nontrivial.

Theorem 8 (Chen and Peng [20]). *Let $U_i \sim \mathcal{N}(0,1)$ $(i = 1, \ldots, k)$ be independent, and let V_1 be the smallest order statistics of n independent standard normal random variables independent of U_i. There exists a constant $\eta > 0$ such that for any $1 \leq k \leq n$,*

$$P\left(\frac{1}{k}\sum_{i=1}^{k} U_i \leq V_1\right) \leq \eta^k \left(\ln\left(\frac{n+k}{k}\right)\right)^{\lceil (k-1)/2 \rceil} \mathrm{B}(k, n+1),$$

where $\mathrm{B}(\cdot,\cdot)$ is the Beta function with

$$\mathrm{B}(n,k) = \int_0^1 u^{n-1}(1-u)^{k-1}\, du = \frac{(n-1)!(k-1)!}{(n+k-1)!}.$$

To bound $P(\mathcal{H}_2^{\mathcal{K}\setminus\{i\}})$, we use an existing result for the GOE from the random matrix theory literature (Dean and Majumdar [28]), which establishes that

$$P(Q \succeq 0) \leq \exp\left(-\frac{n^2}{4}\right).$$

Combining the bounds for $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$ and $P(\mathcal{H}_2^{\mathcal{K}\setminus\{i\}})$ (see Chen and Peng [20] for the detailed derivation and formula), we have the following bound on the probability that a global optimal solution of problem (5) has exactly k nonzero elements.

Theorem 9 (Chen and Peng [20]). *Assume that $Q \in \mathcal{S}^n$ is GOE. Let x^* be one global optimal solution of problem (5). We have that for $k \geq 2$,*

$$P(\|x^*\|_0 = k) \leq \frac{(2k-3)!}{(k-1)!\,(n+1)^{k-2}} \left(\eta^2 \ln\left(\frac{n+k-1}{2k-2}\right)\right)^{k-1} \exp\left(-\frac{(k-2)^2}{4}\right).$$

A careful examination of the right-hand side of the above inequality implies that one can essentially exclude the possibility of global optimal solutions with three or more positive elements for large n. An important implication of Theorem 9 is that if $Q \in \Re^{n\times n}$ is GOE, then with a high probability we can find a global optimal solution of problem (5) by inspecting all feasible solutions with supports no more than two in a polynomial time, though checking the optimality of a feasible solution remains a challenge.

Our result might also lead to a mathematical interpretation of a long observed phenomenon in portfolio selection: the optimal solution of the well-known mean-variance model is dominated by only a few assets (see, for example, Cornuéjols and Tütüncü [24]). A challenge is that the correlation matrix is positive semidefinite, whereas the random matrices analyzed here are indefinite with very high probability.

4. Two-Player Matrix Games

The observation that a StQP may admit a global optimal solution with a support no more than two with a high probability under some conditions carries over to two-player matrix games. Specifically, consider a game with two players 1 and 2. Let $S_1 = \{1, 2, \ldots, m\}$ and $S_2 = \{1, 2, \ldots, n\}$ be the strategy sets of players 1 and 2, respectively. Given player 1's strategy $i \in S_1$ and player 2's strategy $j \in S_2$, players 1 and 2 receive payoffs a_{ij} and b_{ij}, respectively. Let $A = [a_{ij}]_{i=1:\,m,\,j=1:\,n}$ and $B = [b_{ij}]_{i=1:\,m,\,j=1:\,n}$ be the payoff matrices of players 1 and 2. A Nash equilibrium in pure strategies of the resulting two-player matrix game is a pair of pure strategies (i^*, j^*) of both players such that

$$a_{i^*j^*} \geq a_{ij^*} \quad \forall\, i \in S_1 \qquad \text{and} \qquad a_{i^*j^*} \geq a_{i^*j} \quad \forall\, j \in S_2;$$

that is, a pair of pure strategies (i^*, j^*) is a Nash equilibrium if, given player 2's strategy j^*, strategy i^* is player 1's best response, and given player 2's strategy i^*, strategy j^* is player 1's best response. In other words, no player has any incentive to deviate unilaterally from the status quo (i^*, j^*).

A Nash equilibrium in pure strategies may not exist. One remedy is to allow for mixed strategies. Specifically, a mixed strategy for player τ with a strategy set S_τ is a probability distribution defined over S_τ. A mixed strategy pair (μ_1^*, μ_2^*) is a Nash equilibrium (in mixed strategies) if no player has any incentive to deviate unilaterally from the status quo (μ_1^*, μ_2^*), i.e.,

$$u_1(\mu_1^*, \mu_2^*) \geq u_1(i, \mu_2^*) \quad \forall\, i \in S_1 \qquad \text{and} \qquad u_2(\mu_1^*, \mu_2^*) \geq u_2(\mu_1^*, j) \quad \forall\, j \in S_2,$$

where u_1 and u_2 denote the expected payoffs of players 1 and 2, respectively. It is well known that in a matrix game, a Nash equilibrium in mixed strategies always exists. However, the computational complexity of obtaining such a equilibrium in a two-player matrix game was only settled down recently by Chen and Deng [19]. Specifically, they proved that finding a Nash equilibrium (possibly in mixed strategies) is PPAD-complete.

Similar to the motivation of studying the random standard quadratic programs, we would like to understand whether finding a Nash equilibrium would be any easier in some average sense. For this purpose, we consider the following probabilistic model.

Assumption 2. *The random payoffs matrices A and B satisfy the following conditions*:
(a) *A and B are independent.*
(b) *The entries of A is drawn independently from an identical cumulative distribution function $G(\cdot)$.*
(c) *The entries of B is drawn independently from an identical cumulative distribution function $F(\cdot)$.*

Under Assumption 2, Goldberg et al. [34] proved that the probability that there is at least a Nash equilibrium in pure strategies is given by

$$p_{mn} = \sum_{k=1}^{\min(m,n)} (-1)^{k+1} C(n,k) C(m,k) k!\ (mn)^{-k}.$$

They showed that

$$p_{mn} \to 1 - \exp(-1) \quad \text{as } \min(m,n) \to +\infty.$$

Thus, there is a constant probability that a random matrix game does not have a Nash equilibrium in pure strategies.

Focusing on Assumption 2 with uniform distributions on some intervals or Gaussian distributions, Bárány et al. [4] proved that with a high probability, a Nash equilibrium in mixed strategies consists of the mixing of very few pure strategies. Let (μ_1^*, μ_2^*) be a Nash equilibrium in mixed strategies, and let S_τ^* be the support of μ_τ^*, i.e., the set of pure strategies with positive probabilities in μ_τ^*. Bárány et al. [4] first illustrated that given player τ's mixed strategy μ_τ^*, the supports of best response strategies for player κ (κ denotes the player different from τ), S_κ^*, are those supports that induce facets with nonnegative normal vectors in an associated random polytope. Thus, the supports S_1^* and S_2^* have the same cardinality with probability one. By counting the number of faces of any given dimension for a convex hull of n random points, Bárány et al. [4] proved the following theorem (they assume $m = n$ for convenience).

Theorem 10. *The probability that a two-player random matrix game with n strategies for each player contains no Nash equilibrium with support of size at most d is less than*

$$f(d)\left(\frac{1}{n} + \frac{1}{N_1^2}\right),$$

where $f(d)$ is a function of d alone, and N_1 is the expected number of points that lie on the boundary of the convex hull of n random points in d dimensions.

The literature on random polytopes establishes the following results (see Bárány et al. [4] for references): when the n points are drawn independently and uniformly from a d-dimensional unit cube,

$$N_1 \geq g(d)(\log n)^{d-1}$$

for some function $g(d)$ independent of n; when the coordinates of the n points are drawn independently from $\mathcal{N}(0,1)$,

$$N_1 \geq h(d)(\log n)^{(d-1)/2}$$

for some function $h(d)$ independent of n. These results, together with Theorem 10, imply that with a very high probability, a random game has a Nash equilibrium with support of size no more than two, and consequently a Nash equilibrium can be found in a polynomial time with a high probability by exhaustively searching mixed strategies with increasing support cardinalities.

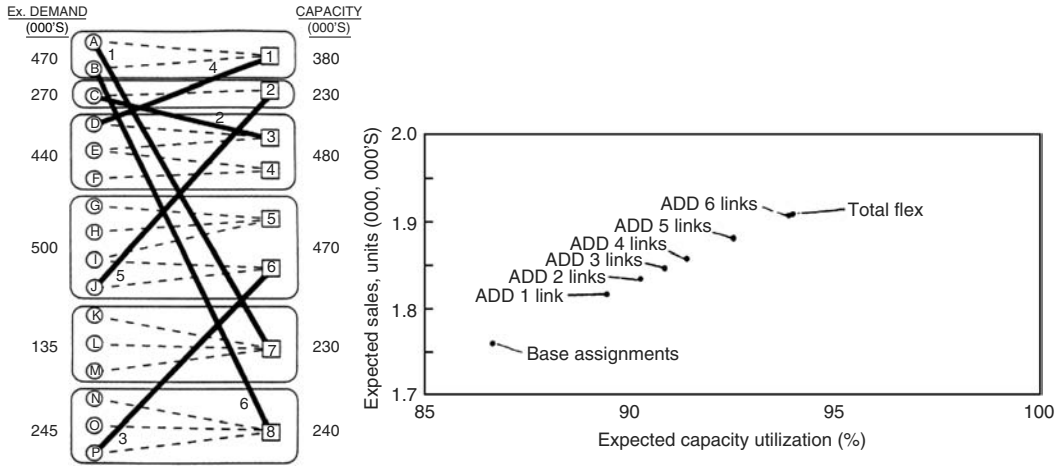
5. Two-Stage Stochastic Program and Flexibility

In this section, we study two-stage stochastic program of the form

$$\max_{z \in \mathcal{C}} \big[E_\eta[g(z,\eta)] - \|z\|_0 \big],$$

where $\mathcal{C}$ is the feasible region for the first-stage decision z, $\|z\|_0$ can be interpreted as the cost of the first-stage decision z, and $g(z,\eta) = \max_{x \in \mathcal{P}(z,\eta)} c(x)$ is the value derived from the

FIGURE 1. Product-plant configuration from a case study in GM.



Source. Jordan and Graves [38].

second-stage decision, with z and η affecting the feasible region $\mathcal{P}(z,\eta)$ for the second-stage decision.

Let z^* denote the optimal solution to this two-stage stochastic program. We are interested in the following question: *Is there a "sparse" solution* $\bar{z}$ *such that* $\|\bar{z}\|_0 \ll \|z^*\|_0$*, whereas* $E_\eta[g(\bar{z},\eta)] \approx E_\eta[g(z^*,\eta)]$?

This question has important ramifications in various settings. In the manufacturing literature, this question is addressed under the general theme of flexible manufacturing system, or, more specifically, the area on "process flexibility." Studies on this area can be traced back to the 1980s, stemming from the hot topic of flexible manufacturing systems (cf. Stecke [51], Browne et al. [13]). The seminal work of Jordan and Graves [38], based on a study of the General Motors (GM) production network, pointed out explicitly that a little flexibility added to the product-plant configuration can enhance the performance of the system (in terms of capacity utilization and product fill rates) substantially. Hence a partial flexibility structure (i.e., a sparse solution) can perform nearly as well as a fully flexible structure (Figure 1).

The recent work by Ambrus et al. [1] on risk-sharing networks formed in Peruvian villages is another instance where the "sparse" solution in tie formation and risk-sharing arrangements can attain close to full insurance. They modelled the social network and the financial and in-kind transfers between relatives and friends in a rural village in the Huaraz province of Peru, and they observed that although risk-sharing arrangements are mainly localized (people helping out mainly neighbors and friends) and thus sparse, the same arrangements often achieve full global insurance at the village level.

It turns out that the performance of the "sparse" solution in these settings is often related to the "expansive" property of the network. To see this, first observe that the problem $E_\eta[g(z,\eta)]$ associated with the process flexibility model boils down to solving the following classic transportation problem on m supply and n demand nodes, with process structure $\mathcal{G} = \{(i,j)\colon z_{ij} = 1\}$:

$$\begin{aligned} g(z,\eta) \equiv \max \ & \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \\ \text{s.t.} \ & \sum_{j=1}^{m} x_{ij} \le \eta_i \quad \forall\, i = 1,2,\ldots,n, \\ & \sum_{i=1}^{n} x_{ij} \le C_j \quad \forall\, j = 1,2,\ldots,m, \end{aligned}$$

$$x_{ij} \geq 0 \quad \forall\, i = 1, \ldots, n,\ j = 1, \ldots, m,$$
$$x_{ij} = 0 \quad \forall\, (i,j) \notin \mathcal{G}.$$

The vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$ encodes the demand for each product, and C_j represents the capacity/supply at plant j. Our goal is to utilize the capacities in the most efficient manner to meet the demands, subject to the constraints imposed by the process flexibility structure $\mathcal{G}$. If η_i's are deterministic, standard LP theory assures that the optimal solution will be sparse, since the basic feasbile solutions have at most $O(n+m)$ nonzero flows. When η_i's are random, clearly the optimal solution z^* will allow flows on as many arcs as possible ($O(mn)$ of them), and the challenge is to find sparse z that can perform nearly as well as z^*.

The problem $E_\eta[g(z,\eta)]$ in the risk-sharing model has similar form:

$$\begin{aligned} g(z,\eta) \equiv \max\ & \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} \\ \text{s.t.}\ & \sum_{j=1}^{n} x_{ij} \leq \eta_i \quad \forall\, i = 1,2,\ldots,n, \\ & \sum_{i=1}^{n} x_{ij} \leq \frac{\sum_i \eta_i}{n} \quad \forall\, j = 1,2,\ldots,n, \\ & x_{ij} \geq 0 \quad \forall\, i = 1,\ldots,n,\ j = 1,\ldots,n, \\ & x_{ij} = 0 \quad \forall\, (i,j) \notin \mathcal{G}, \\ & x_{ij} \leq c_{ij} \quad \forall\, (i,j) \in \mathcal{G}, \end{aligned}$$

where η_i is the endowment shock to i, and c_{ij} is now the strength of the ties between i and j. Note that $n = m$ in this case, and i will transfer an amount t_{ij} to j provided $t_{ij} \leq c_{ij}$, the strength of their ties. We hope to redistribute the wealth within the village to ensure that everyone will have stable consumption close to the average endowments. Again we expect that the risk-sharing network will be most efficient when every pair of individuals in the village can share risk together. Unfortunately, the real social networks often are sparse, and hence the challenge is to explain how sparse social networks can effectively share risk to attain close to full insurance at the village level.

Chou et al. [22] analyzed the performance of a partially flexible structure using a worst case approach. They proved an even stronger result—that a sparse structure is able to perform nearly as well as the fully flexible structure *for all demand realization* within an uncertainty set. To understand this behavior, Chou et al. [22] adopted the concept of graph expansion (cf. Bassalygo and Pinsker [5]), which is widely used in graph theory and computer science (see Sarnak [48] for a good review), to study the process flexibility problem in this setting. Their study reveals the intimate connection between process flexibility and graph connectivity and shows that the graph expander structure (i.e., a class of highly connected graphs with a far smaller number of arcs than a complete graph) works extremely well as a sparse flexibility structure. This result holds in many classes of objective functions, requiring only a mild assumption that the demand is *bounded around its mean*; that is, demand is never more than a constant times its mean.

Definition 2. η_i has a *bounded variation of* λ around its mean if $\eta_i \leq \lambda E[\eta_i]$ almost surely.

Definition 3. A structure $\mathcal{F}$ is *k-connected* if there are at least k node disjoint paths linking *every pair* of nodes in $A \cup B$.

There is a clear trade-off in the level of connectivity with the number of edges—for higher graph connectivity, the structure needs to have more edges. There is a type of highly

connected graph, called a *graph expander*, which has received a lot of attention in the literature. Basically, graph expanders are graphs where every "small" subset of nodes is linked to a "large" neighborhood. The ratio of the size of the neighborhood and the size of the subset measure the graph's expansion capability. We define the neighborhood of a subset and the concept of a graph expander formally in the following.

Definition 4. Let $\mathcal{F}$ be a bipartite graph with partite sets A and B. For $S \subset A$, the *neighborhood of S in $\mathcal{F}$* is defined to be

$$\Gamma(S) = \{j \in B\colon (i,j) \in \mathcal{F} \text{ for some } i \in S\}.$$

Definition 5. Let $\mathcal{F}$ be a bipartite graph with partite sets A and B. The structure $\mathcal{F}$ is an $(\alpha, \lambda, \Delta)$-*expander* if

- for each node v in the graph, $\deg(v) \le \Delta$ for every $v \in A$, and
- for all small subset $S \subset A$ with $|S| \le \alpha n$, we have $|\Gamma(S)| \ge \lambda |S|$.

Since a graph expander ensures that any suitably small group of product nodes is connected to a relatively large number of plants, it works well intuitively in matching supply and demand. Chou et al. [22] established the following main result.

Theorem 11. *Consider an $n \times n$ system, where the demand η_i has a bounded variation of λ with mean $\mu_i = \mu$. Assume that each plant has capacity μ. Let z gives rise to the structure $\mathcal{G}$, which is an $(\alpha, \lambda, \Delta)$-expander, with $\alpha \times \lambda = 1 - \epsilon$ for some $\epsilon > 0$. Then,*

$$g(z,\eta) \ge \alpha\lambda n \left[\min\left(\mu, \frac{\sum_{i\in A}\eta_i}{n}\right)\right] = (1-\epsilon) g(z^*, \eta)$$

for all η_i satisfying the assumptions.

To find a good process structure for a large n using only a small number of links, we use an expander where the number of edges can be *much smaller* than the number of edges in a fully flexible system. The existence of such a structure is well known and is by now folklore in the graph theory community.

Theorem 12 (Asratian et al. [2]). *For any n, $\lambda \ge 1$, and $\alpha < 1$ with $\alpha\lambda < 1$, there exists an $(\alpha, \lambda, \Delta)$-expander, for any*

$$\Delta \ge \frac{1 + \log_2 \lambda + (\lambda+1)\log_2 e}{-\log_2(\alpha\lambda)} + \lambda + 1. \tag{19}$$

Note that the lower bound on degree Δ is independent of n, and hence the number of edges in this class of graph expanders is *linear* in n. The implication for the design of a flexible process structure can be stated more succinctly as follows: *In a symmetrical system, for any given demand distribution with a bounded variation of λ, we can find a corresponding α with*

$$\alpha\lambda \approx 1 - \epsilon, \quad \text{for some } \epsilon > 0,$$

such that for a sufficiently large n, we can always find a process structure using at most Δn edges, where Δ is given by the right-hand side of (19)*, such that the performance of the structure is at most $1 - \epsilon$ times that of the fully flexible system even in the worst-case scenario.*

For more general systems (i.e., the number of product nodes and plant nodes might differ and products might follow different demand distributions), Chou et al. [22] proposed a generalization using the concept of "Ψ-expander" with a high Ψ $(0 < \Psi \le 1)$. Suppose the demand with mean μ_i is bounded in $[\lambda_i^L \mu_i, \lambda_i^U \mu_i]$. We say that the demand has a bounded variation of λ_i^L and λ_i^U in this case.

Definition 6. Given Ψ, where $0 < \Psi \leq 1$, a Ψ-*expander* in the process flexibility problem is a bipartite graph in $A \times B$ with

$$\sum_{j \in \Gamma(S)} C_j \geq \min\left\{ \sum_{i \in S} \lambda_i^U \mu_i, \Psi \sum_{j \in \mathcal{B}} C_j - \sum_{i \notin S} \lambda_i^L \mu_i \right\}$$

for all subsets $S \subseteq A$.

Note that every network is at least a 0-expander. The challenge in practice is to find the largest Ψ such that the above holds.

The definition of a Ψ-expander partitions the subsets of A into two groups, *small* and *nonsmall* subsets.

Definition 7. Given a Ψ-expander, we refer to a subset $S \subseteq A$ as a *small* subset if

$$\sum_{i \in S} \lambda_i^U \mu_i \leq \Psi \sum_{j \in \mathcal{B}} C_j - \sum_{i \notin S} \lambda_i^L \mu_i.$$

For any $S \subseteq A$ that is not a *small* subset, we call it a *nonsmall* subset.

- For *small* subset S, we have

$$\sum_{j \in \Gamma(S)} C_j \geq \sum_{i \in S} \lambda_i^U \mu_i,$$

and hence the plants supplying it have sufficient capacity to deal with the demand arising from it.

- At the same time, the capacity connected to a *nonsmall* subset is also large enough; that is,

$$\sum_{j \in \Gamma(S)} C_j \geq \Psi \sum_{j \in B} C_j - \sum_{i \notin S} \lambda_i^L \mu_i,$$

so that at least Ψ proportion of the total capacity is utilized in the worst case. It is thus easy to see that a structure with $\Psi = 1$ is as good as full flexibility, and the larger Ψ is, the more flexible is a structure.

A more intuitive way to understand the above is to consider the special case when $\lambda_i^L = 0$ for all i. Our definition for a Ψ-expander reduces to

$$\sum_{j \in \Gamma(S)} C_j \geq \min\left\{ \sum_{i \in S} \lambda_i^U \mu_i, \Psi \sum_{j \in \mathcal{B}} C_j \right\} \quad \forall\, S \subseteq A.$$

For a 1-expander, we need

$$\sum_{j \in \Gamma(S)} C_j \geq \sum_{i \in S} \lambda_i^U \mu_i \quad \forall\, S \subseteq A.$$

From the max-flow min-cut theorem, this simply means that the capacities embedded in the system must be able to support a maximum flow of $\sum_{i \in A} \lambda_i^U \mu_i$. This is no doubt too strong a requirement. By choosing $\Psi < 1$, we relax the cut conditions considerably but impose a new condition that at least $100 \times \Psi\%$ of the capacity must be utilized when total demand is large.

Theorem 13 (Chou et al. [22]). *Let $\mathcal{F}$ be a Ψ-expander. When η_i has a bounded variation of λ_i^L and λ_i^U, then for all demand realizations, we can find a solution for $Z_{\mathcal{F}}$ such that either* (a) *all the plants are operating below their configured capacity level* (*because of insufficient demand*) *or* (b) *at least Ψ proportion of the total preconfigured capacity has been utilized.*

Table 1. Expected sales and chaining efficiency for increasing system size.

System size n	Chaining $\mathcal{C}(n)$	Full flexibility $\mathcal{F}(n)$	Ratio of max-flow between chain and full (%)
10	949.36	955.14	99.39
15	1,434.44	1,447.00	99.13
20	1,915.78	1,938.93	98.81
25	2,401.94	2,441.73	98.37
30	2,871.06	2,929.84	97.99
35	3,352.66	3,430.70	97.73
40	3,807.16	3,905.48	97.48

The above theorem suggests that a Ψ-expander has the following nice property: as long as the demand for each product falls in the range of $\lambda_i^L \mu_i$ and $\lambda_i^U \mu_i$, then the process structure guarantees a utilization rate of $100 \times \Psi\%$ in the entire system.

When the demand distribution is explicitly modelled, analyzing the second-stage stochastic program in the process flexibility problem is a lot harder, because it reduces to the classical stochastic maximum flow problem. In the identical and balanced case, it is already known from Jordan and Graves [38] that a two-chain works exceedingly well. Chou et al. [23] use a random walk model to explicitly characterize the asymptotic performance of the two-chain for a variety of distributions, whereas Wei and Simchi-Levi [55] use the classic results on the comparative statics of maximum flow on circulation to strengthen the analysis. Two (directed) arcs are said to be "parallel" if every (undirected) simple cycle containing both of them orients them in the opposite direction, and "series" if every (undirected) simple cycle containing both of them orients them in the same direction. A set of arcs is "parallel" if all pairs of arcs are parallel, and "series" if all pairs are series arcs.

Proposition 2 (Gale and Politof [31]). *Let P be a "parallel" arc set and S a "series" arc set in $\mathcal{G}$. Then the maximum flow $E_\eta[g(z,\eta)]$ is submodular in P and supermodular in S.*

It follows that

Theorem 14 (Wei and Simchi-Levi [55]). *In a balanced and symmetrical system, and among all 2-regular configurations with $2n$ arcs, the 2-chain has the highest expected maximum flow.*

Chou et al. [23] demonstrated this effect more succinctly by comparing the performance of the chaining structure with the fully flexible system for an asymptotically large system (balanced and identical). When the demand is uniformly distributed between 0 and $2C$,

$$\lim_{n\to\infty} \frac{\text{expected max-flow on 2-chain}}{\text{expected max-flow on complete graph}} = 89.6\%.$$

This implies that a simple chain structure can capture close to 90% of the value of the maximum flow in a fully flexible system, *even when the system size is very large* and the demand is uniform over a range. The performance in the case of normal distribution is even more impressive. Table 1 shows the expected performance of two different structures over the random demand as n varies, assuming that $\mu = 3\sigma$.

As n approaches infinity, the limit of the ratio tends to a value close to 96% (cf. Chou et al. [23]).

6. Applications of Sparse Solution

The central theme in this tutorial is the observation that a sparse solution can go a long way in enhancing the performance of a system. We have discussed the impact of this phenomenon on the process flexibility problem in the earlier sections. The effectiveness of risk-sharing

performance on real social network can also be attributed to the expansive property of the social network. In the rest of this section, we review some of the key results obtained for other related areas and discuss some new applications.

6.1. Military Deployment and Transshipment

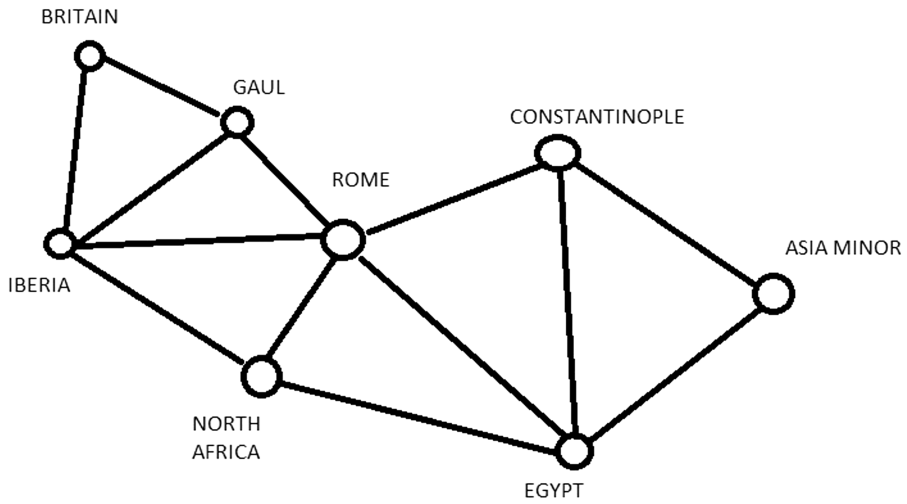
Inspired by the defense-in-depth strategy devised by Emperor Constantine (Constantine the Great, ca. 272–337), ReVelle and Rosing [46] studied the following problem in troop deployment: Each region in the empire must be protected by one or more mobile field armies (FAs) to throw back invading enemies. It is secured if one or more FAs are stationed in the region. It is securable if an FA can reach the region in a single step (i.e., there is a route linking the region to where the FA is stationed). However, an FA can be deployed from one region to an adjacent region only when there is at least one other FA to help launch it; that is, the FA must come from a region that has at least two FAs stationed in it. This restriction is much like the island-hopping strategy used by General MacArthur in World War II in the Pacific.

The puzzle confronting Emperor Constantine concerned the positioning of four FAs to protect the eight regions in his empire, as shown in Figure 2. He chose to position two FAs at Rome and two at his new capital in Constantinople, leaving the outskirt region of Britain vulnerable to enemy attack. By focusing on the troop deployment problem in the event of war in one of the regions, ReVelle and Rosing [46] solved the above puzzle by formulating the problem into an integer program. In this case, all regions in the empire can be protected by stationing one FA in Britain, one in Asia Minor, and two in Rome.

In general, finding the best deployment solution securing against an outbreak of wars in up to k regions (for $k \geq 2$) is a challenging problem. The minimum number of FAs needed to secure the regions will largely depend on the network structure for troop redeployment. In general, if the network is dense (with many links joining different regions) or has one region connecting to many different regions, then the number of FAs needed will be low.

In this section, we consider an analogous military deployment problem. Consider a military mission where n strategic locations need to be defended against a possible enemy invasion. The army has Q_i units of troops in location i. Unfortunately, the enemy's mission cannot be predicted. The unit of troops deployed by the enemy to attack location i is denoted by D_i. One way to strengthen the defense network is to have reinforcement troops whereby units in location i may be deployed to location j if the troops can be trained to rush from i to j

Figure 2. The empire of Constantine.



within a stipulated time. Of course, it would be ideal to have many reinforcement paths, as that would mean the whole force could be pooled together at the right place to deal with the enemy's invasion. But because of the limited time in deployment, each unit in location i can only be trained to reinforce a limited number of other locations. The challenge is to design a reinforcement network to defend against the enormous number of the enemy's possible courses of action.

This problem is similar to the transshipment problem studied in the literature, although the latter focuses mainly on the optimal inventory policy and optimal order quantity Q_i^* for each retailer (for problems with two retailers, see Tagaras and Cohen [52]; for problems with many identical retailers, see Robinson [47]). Most of the studies on the transshipment problem assume a complete grouping; that is, a retailer could transship its products to any other retailer. Only a few papers have discussed how to design a transshipment network. Lien et al. [41] studied the differing impacts of a transshipment network structure by comparing the performance of different network configurations. Similar to findings in Jordan and Graves [38], they showed that a sparse transshipment network structure can capture almost all the benefits of a complete grouping. They also indicated that the chaining structure, which is also a kind of sparse structure, would outperform other sparse structures.

The troop deployment (and the transshipment) problem can be reduced to a variant of the process flexibility problem, where there are n plants and n products. Each plant i has capacity $(Q_i - D_i)^+$ (the leftover at retailer i), which can be used to meet the demand for other products. Each product has demand $(D_i - Q_i)^+$ (unfilled demand at retailer i). Note that in this case, both capacity and demand are random parameters in our problem, and $(Q_i - D_i)^+ \times (D_i - Q_i)^+ = 0$.

The existence of a sparse support structure for the troop deployment problem is guaranteed by the following condition:

$$\begin{aligned} x_{i,j}^*(D) &= \frac{(D_i - Q_i)^+ (Q_j - D_j)^+}{\max\{\sum_{i=1}^n (D_i - Q_i)^+, \sum_{j=1}^n (Q_j - D_j)^+\}} \\ &\le \lambda E_D \left[\frac{(D_i - Q_i)^+ (Q_j - D_j)^+}{\max\{\sum_{i=1}^n (D_i - Q_i)^+, \sum_{j=1}^n (Q_j - D_j)^+\}} \right] \end{aligned}$$

almost surely for some $\lambda > 1$ and for all i, j.

There is a combinatorial analogue to the troop deployment problem. Suppose we distribute $2n$ units of troops uniformly on $2n$ nodes, with each location defended by exactly one unit. Suppose also that each location will not be penetrable only if two units are defending that location. If the enemy can attack up to n different locations, how would we design the reinforcement network?

On the other hand, if the enemy chooses to attack without knowing the reinforcement network, the problem can be reduced to the random allocation of n red and n blue balls uniformly in the nodes of the network. Let $c(i)$ denote the color assigned to node i. Let $E(G)$ denote the edge set in G. We say that $M \subset E(G)$ is a colored matching if it is a matching in G with

$$M = \big\{(i,j)\colon c(i) \ne c(j), (i,j) \in E(G)\big\}.$$

Let $m(G)$ denote the cardinality of a maximum colored matching in G. Thus, $m(G)$ represents the number of locations that can be defended in the network. Note that $m(G) \le n$ for all realizations of the color distribution, and $E(m(G)) = n$ when $E(G) = K(2n)$, the complete graph on $2n$ nodes.

The cardinality of the edge set $E(G)$ can be reduced much further, while sacrificing only a little of the value of $E(m(G))$.

Theorem 15. *For all $\epsilon > 0$ there exists $n(\epsilon) > 0$ such that for all $n \ge n(\epsilon)$ there exists a graph G_n with $2n$ nodes and $O(n)$ edges, such that*

$$n \ge E(m(G_n)) \ge (1 - \epsilon)n.$$

Hence, a sparse but near-to-optimal reinforcement network can be obtained with only a small loss of locations.

6.2. Load Balancing

The concept of limited flexibility also has important applications in load balancing in stochastic network routing analysis (cf. Mitzenmacher [43]). This application follows from the following interesting observation: suppose m balls are randomly inserted into n bins, with each bin chosen with probability $1/n$. What is the expected number of balls in a bin with the maximum load?

Since the expected number of balls per bin is m/n, if m is much larger than n, say $m \geq 2n\log_2(n)$, then we expect the number of balls per bin to concentrate around this mean. In fact, one can show that (cf. Mitzenmacher [43] and the references therein).

Theorem 16. *Suppose $m \geq 2n\log_2(n)$. With high probability, i.e., $1-1/n$, all bins have at most $e(m/n)$ balls.*

In the light load case, $m < 2n\log_2 n$, say $m = n$, it is actually not difficult to show that the bin with the maximum load should have $O(\log(n))$ balls with a high probability. This is surprising because the mean load is $O(1)$ in the case $m = n$, but the maximum load is significantly higher than the mean with high probability.

Theorem 17. *For $m < 2n\log_2 n$, with high probability, i.e., $1-1/n$, all bins have at most $(4e^2\ln(n))/(\ln((2en/m)\ln(n)))$ balls.*

Suppose we modify the process in the following way: the balls are inserted into the bins sequentially. Each ball gets to pick k bins randomly. Depending on the load on the k bins at that time, the ball will be inserted into the bin with the smaller load. Ties are resolved arbitrarily.

Theorem 18 (Mitzenmacher [43]). *For k random choices, the maximum load is*

$$\frac{m}{n} + \frac{\log\log(n)}{\log k} + O(1)$$

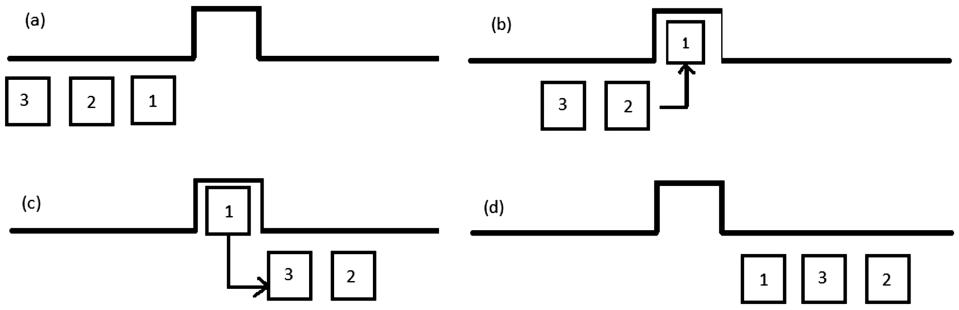
with high probability.

It turns out that this simple modification reduces the peak load drastically to $O(\log\log n)$ with a high probability. Having more flexibility does not help much either, since if we allow each ball to pick K bins randomly, then the peak load is reduced to $O((\log\log n)/\log K)$, for any $K \geq 2$; that is, having more flexibility only reduces the peak load by a constant factor.

6.3. Sequencing with Limited Flexibility

Lahmar et al. [39] considered the following sequencing problem in an automotive assembly line: In the paint area, vehicles undergo a series of painting operations. If two consecutive vehicles are painted different colors, a significant changeover cost is incurred since the current paint must be flushed out and disposed of, and paint nozzles must be thoroughly washed and cleaned with solvents. Hence cars leaving the body shop on a moving line must be resequenced prior to entering the paint shop to minimize the changeover costs at the paint shop. Boysen et al. [12, p. 277] provided a good overview of this class of problems: "Common and widespread forms of resequencing buffers in the automobile industry are selectivity banks and pull-off tables. Selectivity banks consist of a set of parallel first-in-first-out lanes. Models are assigned to one of the lanes, enter the lane on, e.g., the left-hand side and move forward to the right-hand side. Only models on the right-hand side of each lane are accessible to proceed downstream. Thus, the number of models to choose from is bounded by the number of lanes. In contrast, pull-off tables (see Figure 3) are direct accessible buffers.

FIGURE 3. Pull-off table in automobile production.



A model in the sequence can be pulled into a free pull-off table, so that successive models can be brought forward and processed before the model is reinserted from the pull-off table back into a later sequence position."

The resequencing problem with position shifting constraints can be stated as follows: Given an initial ordering σ of n jobs, find a minimum cost permutation π of σ, satisfying (i) $\pi(i)+K_1 \geq i$ and (ii) $\pi(i)-K_2 \leq i$ for all i, where π is defined as a one-to-one mapping from each job (denoted by its position in σ) to its position in the final ordering such that $\pi(i)$ represents the position of job i in the final sequence, and K_1, K_2 are positive integers.

Given an initial ordering of jobs, they proposed a dynamic program (DP) to find the minimum cost permutation of the sequence, so that each position is shifted not more than K_1 positions to the right, and not more than K_2 positions to the left. The values K_1 and K_2 reflect the limited buffer space available in the production plant, as well as the level of flexibility within the plant. Resequencing is needed to minimize, say, the changeover costs at the next station. A precise analytical measurement of the value of flexibility, however, is difficult to obtain, because the complexity of the DP-based algorithm depends on the values of K_1 and K_2. Nevertheless, the numerical results in this paper are quite convincing: the effect of flexibility diminishes rapidly, and most of the benefits can be accrued at small values of K_1 and K_2. Lahmar et al. [39] shows an experiment with $K_2 = n-1$, and as K_1 varies, the benefits of flexibility diminishes rapidly for different changeover cost distributions.

Nevertheless, finding a theory to explain this phenomenon remains an outstanding open problem. One way of measuring sequencing flexibility is to count the total number of feasible sequences that can be produced. It turns out that this counting problem has a rich history in combinatorics and is recently used in the area of steganographic communication in ordered channels, where ordered packets are re-sequenced to hide the information in the transmission. The reordering is usually done by either distance-bounded permutation (to ensure a bound on the latency of the transmission) or buffer-bounded permuter. This problem is studied by Lehmer [40]: Let $N(n,k,r)$ denote the number of strongly restricted permutations of $[1,n]$ satisfying the conditions $-k \leq \pi(i)-i \leq r$ (for arbitrary natural numbers k and r).

The class of permutations in which the positions of the marks after the permutation are restricted can be specified by a matrix $A=(a_{ij})$ in which $a_{ij}=1$ if the mark j is permitted to occupy the ith place and 0 otherwise.

Proposition 3. *The number of restricted permutations is given by the permanent function of a square matrix A:*

$$\operatorname{per}(A) = \sum_{p \in S_n} a_{1p(1)} a_{2p(2)} \dots a_{np(n)},$$

where p runs through the set S_n of all permutations of $[1,n]$.

Remarkably, Vladimir [53] showed recently that $N(n,1,r)$ is simply the Fibonacci $(r+1)$-step number.

Example 1. The 2-step Fibonacci numbers are the well-known series

$$\underbrace{1}_{0}, \underbrace{1}_{1}, \underbrace{2}_{2}, \underbrace{3}_{3}, \underbrace{5}_{4}, \underbrace{8}_{5}, \underbrace{13}_{6}, \underbrace{21}_{7}, \ldots,$$

so that $N(3,1,1) = 3, N(4,1,1) = 5$, etc. The 3-step Fibonacci numbers are

$$\underbrace{1}_{0}, \underbrace{1}_{1}, \underbrace{2}_{2}, \underbrace{4}_{3}, \underbrace{7}_{4}, \underbrace{13}_{5}, \underbrace{24}_{6}, \underbrace{44}_{7}, \ldots,$$

so $N(3,1,2) = 4, N(4,1,2) = 7$, etc. The 4-step Fibonacci numbers are

$$1, 1, 2, 4, 8, 15, 29, 56, 108, \ldots,$$

so $N(4,1,3) = 8$.

In general, computing the values of $N(n,k,r)$ is still an outstanding problem. Benjaafar and Ramakrishnan [8] presented a comprehensive review of other viable measurements for sequencing flexibility and performed a thorough simulation to compare the performance of different measurements.

6.4. Other Applications

The flexibility strategy has been shown to be rather effective in various other areas such as supply chain planning (Bish and Wang [9], Bish et al. [10]), queuing (Bassamboo et al. [6], Benjafaar [7], Gurumurthi and Benjaafar [35]), revenue management (Gallego and Phillips [32]), scheduling (Daniels and Mazzola [25], Daniels et al. [26, 27]), and flexible work force scheduling (Hopp et al. [36], Whitt and Wallace [54], Brusco and Johns [14]). For instance, Hopp et al. [36] observed similar results in their study of a work force scheduling problem in a constant work-in-process queuing system. By comparing the performances of "cherry-picking" and "skill-chaining" cross-training strategies, they observed that "skill chaining," which is indeed a kind of "chaining" strategy, outperformed the other strategies. They also showed that a chain with a low degree (the number of tasks a worker can handle) is able to capture the bulk of the contribution from a chain with a high degree. Whitt and Wallace [54] explored the use of chaining in call center staffing and skill chaining. They showed that, in a scenario where the duration of service does not depend on the call type or the agents serving it, a simple routing policy, together with proper skill chaining, can result in near-optimal performance, even when taking service level constraints (e.g., a service level guarantee for type k calls, or bounds on blocking probability) into consideration. This paper shows that it may be more worthwhile to pay attention to cross-training, rather than investing in complicated call-routing software. The proper staffing levels are then identified via simulation-based optimization.

Acknowledgments

The first author was partly supported by National Science Foundation [Grants CMMI-0653909, CMMI-0926845 ARRA, and CMMI-1030923] and China NSFC [Grant 71228203]. He would like to thank Professor Chung Piaw Teo and the department of decision sciences at National University of Singapore for hosting his visit during which this tutorial was prepared. The authors also want to thank Jiawei Zhang for valuable comments and suggestions.

References

[1] A. Ambrus, M. Mobius, and A. Szeidl. Consumption risk-sharing in social networks. *American Economic Review*. Forthcoming, 2013.

[2] A. Asratian, T. Denley, and R. Haggkvist. *Bipartite Graphs and Their Applications*. Cambridge University Press, Cambridge, UK, 1998.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation* 28(3):253–263, 2008.

[4] I. Bárány, S. Vempala, and A. Vetta. Nash equilibria in random games. *Random Structures and Algorithms* 31(4):391–405, 2007.

[5] L. A. Bassalygo and M. S. Pinsker. Complexity of an optimum non-blocking switching network with reconnections. *Problemy Informatsii* (*English Translation in Problems of Information Transmission*) 9:84–87, 1973.

[6] A. Bassamboo, R. S. Randhawa, and J. A. Van Mieghem. A little flexibility is all you need: Optimality of tailored chaining and pairing. *Operations Research* 60(6):1423–1435, 2012.

[7] S. Benjafaar. Modeling and analysis of congestion in the design of facility layouts. *Management Science* 48(5):679–704, 2002.

[8] S. Benjaafar and R. Ramakrishnan. Modeling, measurement and evaluation of sequencing flexibility in manufacturing systems. *International Journal of Production Research* 34(5):1195–1220, 1996.

[9] E. Bish and Q. Wang. Optimal investment strategies for flexible resources, considering pricing and correlated demands. *Operations Research* 52(6):954–964, 2004.

[10] E. Bish, A. Muriel, and S. Biller. Managing flexible capacity in a make-to-order environment. *Management Science* 51(2):167–180, 2005.

[11] I. M. Bomze, M. Locatelli, and F. Tardella. New and old bounds for standard quadratic optimization: Dominance, equivalence and incomparability. *Mathematical Programming* 115(1):31–64, 2008.

[12] N. Boysen, U. Golle, and F. Rothlauf. The car resequencing problem with pull-off tables. *Business Research* 4(2):276–292, 2011.

[13] J. Browne, D. Dubois, K. Rathmill, S. P. Sethi, and K. E. Stecke. Classification of flexible manufacturing systems. *FMS Magazine* 2(2):114–117, 1984.

[14] M. Brusco and T. Johns. Staffing a multiskilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Science* 29(2):499–515, 1998.

[15] J. A. Buzacott and M. Mandelbaum. Flexibility in manufacturing and services: Achievements, insights and challenges. *Flexible Services and Manufacturing Journal* 20(1–2):13–58, 2008.

[16] E. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique* 346(9–10):589–592, 2008.

[17] E. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions of Information Theory* 51(12):4203–4215, 2005.

[18] S. S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* 20(1):33–61, 1998.

[19] X. Chen and X. Deng. Settling the complexity of 2-player Nash-equilibrium. *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, CA, 2006.

[20] X. Chen and J. Peng. New analysis on sparse solutions to random standard quadratic optimization problems and extensions. Technical report, University of Illinois at Urbana–Champaign, Urbana, 2012.

[21] X. Chen, J. Peng, and S. Zhang. Sparse solutions to random standard quadratic optimization problems. *Mathematical Programming*. Forthcoming, 2013.

[22] M. C. Chou, C. P. Teo, and H. Zheng. Process flexibility revisited: The graph expander and its applications. *Operations Research* 59(5):1090–1105, 2011.

[23] M. C. Chou, G. Chua, C. P. Teo, and H. Zheng. Design for process flexibility: Efficiency of the long chain and sparse structure. *Operations Research* 58(1):43–58, 2010.

[24] G. Cornuéjols and R. Tütüncü. Optimization methods in finance. Cambridge University Press, Cambridge, UK, 2007.

[25] R. L. Daniels and J. B. Mazzola. Flow shop scheduling with resource flexibility. *Operations Research* 42(3):504–522, 1994.

[26] R. L. Daniels, B. J. Hoopes, and J. B. Mazzola. Scheduling parallel manufacturing cells with resource flexibility. *Management Science* 42(9):1260–1276, 1996.

[27] R. L. Daniels, J. B. Mazzola, and D. Shi. Flow shop scheduling with partial resource flexibility. *Management Science* 50(5):658–669, 2004.

[28] D. S. Dean and S. N. Majumdar. Extreme value statistics of eigenvalues of Gaussian random matrices. *Physical Review E* 77:041108.1–041108.12, 2008.

[29] D. Donoho. For most large underdetermined systems of linear equations, the minimal l_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* 59(6):797–829, 2006.

[30] Y. C. Eldar and G. Kutyniok, editors. *Compressed Sensing: Theory and Applications.* Cambridge University Press, Cambridge, UK, 2012.

[31] D. Gale and T. Politof. Substitutes and complements in networks flow problems. *Discrete Applied Mathematics* 3(3):175–186, 1981.

[32] G. Gallego and R. Phillips. Revenue management of flexible products. *Manufacturing and Service Operations Management* 6(4):321–337, 2004.

[33] L. E. Gibbons, D. W. Hearn, P. Pardalos, and M. V. Raman. Continuous characterizations of the maximal clique problem. *Mathematics of Operations Research* 22(3):754–768, 1997.

[34] K. Goldberg, A. Goldman, and M. Newman. The probability of an equilibrium point. *Journal of Research of the National Bureau of Standards* 72(2):93–101, 1968.

[35] S. Gurumurthi and S. Benjaafar. Modeling and analysis of flexible queueing systems. *Naval Research Logistics* 51(5):755–782, 2004.

[36] W. J. Hopp, E. Tekin, and M. P. Van Oyen. Benefits of skill chaining in production lines with cross-trained workers. *Management Science* 50(1):83–98, 2004.

[37] T. Irabaki and N. Katoh. *Resource Allocation Problems: Algorithmic Approaches.* MIT Press, Cambridge, MA, 1988.

[38] W. C. Jordan and S. C. Graves. Principles on the benefits of manufacturing process flexibility. *Management Science* 41(4):577–594, 1995.

[39] M. Lahmar, H. Ergan, and S. Benjaafar. Resequencing and feature assignment on an automated assembly line. *IEEE Transactions on Robotics and Automation* 19(1):89–102, 2003.

[40] D. H. Lehmer. Permutations with strongly restricted displacements. P. Erdös, A. Renyi, V. T. Sós, eds. *Combinatorial Theory and Its Applications II*, North-Holland, Amsterdam, 755–770, 1970.

[41] R. Lien, S. M. Iravani, K. Smilowitz, and M. Tzur. Efficient and robust design for transshipment networks. *Production and Operations Management* 20(5):699–713, 2011.

[42] H. M. Markowitz. Portfolio selection. *Journal of Finance* 7(1):77–91, 1952.

[43] M. Mitzenmacher. The power of two choices in randomized load balancing. Ph.D. dissertation, University of California, Berkeley, Berkeley, 1996.

[44] E. S. Olcott. Innovative approaches to urban transportation planning. *Public Administration Review* 33(3):215–224, 1973.

[45] M. Phelan. Ford speeds changeovers in engine production. *Knight Rider Tribune Business News* (November 6), 2002.

[46] C. S. ReVelle and K. E. Rosing. Defendens imperium romanum: A classical problem in military strategy. *American Mathematical Monthly* 107(7):585–594, 2000.

[47] L. W. Robinson. Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. *Operations Research* 38(2):278–295.

[48] P. Sarnak. What is an expander? *Notices of the American Mathematical Society* 51(7):762–763, 2004.

[49] A. Scozzari and F. Tardella. A clique algorithm for standard quadratic programming. *Discrete Applied Mathematics* 156(13):2439–2448, 2008.

[50] A. K. Sethi and S. P. Sethi. Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems* 2(4):289–328, 1990.

[51] K. E. Stecke. Formulation and solution of nonlinear integer production planning problems for flexible manufacturing systems. *Management Science* 29(3):273–288, 1983.

[52] G. Tagaras and M. Cohen. Pooling in two-location inventory systems with non-negligible replenishment lead times. *Management Science* 38(8):1067–1083, 1992.

[53] B. Vladimir. On the number of certain types of strongly restricted permutations. *Applicable Analysis and Discrete Mathematics* 4(1):119–135, 2010.

[54] W. Whitt and R. B. Wallace. A staffing algorithm for call centers with skill-based routing. *Manufacturing and Service Operations Management* 7(4):276–294, 2005.

[55] Y. Wei and D. Simchi-Levi. Understanding the performance of the long chain and sparse designs in process flexibility. *Operations Research* 60(5):1125–1141, 2012.

[56] S. Yang and X. Li. Algorithms for determining the copositivity of a given matrix. *Linear Algebra and Its Applications* 430(2–3):609–618, 2009.