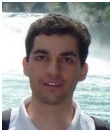
Convex sets, conic matrix factorizations and conic rank lower bounds

Pablo A. Parrilo

Laboratory for Information and Decision Systems
Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Based on joint work with **João Gouveia** (U. Coimbra), **Rekha Thomas** (U. Washington), and **Hamza Fawzi** (MIT)



Nonnegative factorizations

Given a nonnegative matrix $A \in \mathbb{R}^{n \times m}$, a factorization

$$A = UV$$

where $U \in \mathbb{R}^{n \times k}$, $V \in \mathbb{R}^{k \times m}$ are also nonnegative.

- The smallest such k is the *nonnegative rank* of the matrix A.
- Many applications: statistics, factor models, machine learning, ...
- Very difficult problem, many heuristics exist.

Factorizations and hidden variables

Let X, Y be discrete random variables, with joint distribution

$$\mathbf{P}[X = i, Y = j] = P_{ij}.$$

The nonnegative rank of P is the smallest support of a random variable W, such that X and Y are conditionally independent given W (i.e., $X - W - Y$ is Markov):

$$\mathbf{P}[X = i, Y = j] = \sum_{s=1,\ldots,k} \mathbf{P}[X = i, Z = s] \cdot \mathbf{P}[Y = j, Z = s].$$

- Relations with information theory, "correlation generation," communication complexity, etc.
- Quantum versions are also of interest.

As we'll see, fundamental in optimization …

Factorizations and hidden variables

Let X, Y be discrete random variables, with joint distribution

$$\mathbf{P}[X = i, Y = j] = P_{ij}.$$

The nonnegative rank of P is the smallest support of a random variable W, such that X and Y are conditionally independent given W (i.e., $X - W - Y$ is Markov):

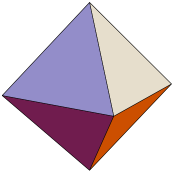
$$\mathbf{P}[X = i, Y = j] = \sum_{s=1,\ldots,k} \mathbf{P}[X = i, Z = s] \cdot \mathbf{P}[Y = j, Z = s].$$

- Relations with information theory, "correlation generation," communication complexity, etc.
- Quantum versions are also of interest.

As we'll see, fundamental in optimization . . .

Motivating example

The *crosspolytope* $\mathcal{C}_n$ is the unit ball of the ℓ_1 ball:



$$\mathcal{C}_n := \{x \in \mathbb{R}^n : \sum_{i=1}^{n} |x_i| \leq 1\}.$$

It is a polytope defined by 2^n linear inequalities:

$$\pm x_1 \pm x_2 \pm \cdots \pm x_n \leq 1$$

The "obvious" linear program is exponentially large!

A better representation

By introducing *slack* or *auxiliary* variables, the set $\mathcal{C}_n$ can be represented more conveniently:

$$\mathcal{C}_n := \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^n, \quad -y_i \leq x_i \leq y_i, \quad \sum_{i=1}^{n} y_i = 1\}.$$

This has only $2n$ variables $(x_1, y_1, \ldots, x_n, y_n)$ and $2n + 1$ constraints. A “small” linear program. Much better!

What is going on in here?

A better representation

By introducing *slack* or *auxiliary* variables, the set $\mathcal{C}_n$ can be represented more conveniently:

$$\mathcal{C}_n := \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^n, \quad -y_i \leq x_i \leq y_i, \quad \sum_{i=1}^{n} y_i = 1\}.$$

This has only $2n$ variables $(x_1, y_1, \ldots, x_n, y_n)$ and $2n + 1$ constraints. A "small" linear program. Much better!

What is going on in here?

Geometric viewpoint

Geometrically, we are representing our polytope as a *projection* of a higher-dimensional polytope.

The number of *vertices* does not increase, but the number of *facets* can grow exponentially!

“Complicated” objects are sometimes easily described as “projections” of “simpler” ones.

A general theme: algebraic varieties, graphical models, . . .

Geometric viewpoint

Geometrically, we are representing our polytope as a *projection* of a higher-dimensional polytope.

The number of *vertices* does not increase, but the number of *facets* can grow exponentially!

"Complicated" objects are sometimes easily described as "projections" of "simpler" ones.

A general theme: algebraic varieties, graphical models, ...

Extended formulations

These representations are usually called *extended formulations*.
Particularly relevant in combinatorial optimization (e.g., TSP).

Seminal work by Yannakakis (1991), who used them to disprove the existence of a "symmetric" LP formulation for the TSP polytope.
Nice recent survey by Conforti-Cornuéjols-Zambelli (2010).

Our goal: to understand this phenomenon for convex optimization, not just LP.

Extended formulations

These representations are usually called *extended formulations*.
Particularly relevant in combinatorial optimization (e.g., TSP).

Seminal work by Yannakakis (1991), who used them to disprove the existence of a “symmetric” LP formulation for the TSP polytope.
Nice recent survey by Conforti-Cornuéjols-Zambelli (2010).

Our goal: to understand this phenomenon for convex optimization, not just LP.

Extended formulations

These representations are usually called *extended formulations*.
Particularly relevant in combinatorial optimization (e.g., TSP).

Seminal work by Yannakakis (1991), who used them to disprove the existence of a "symmetric" LP formulation for the TSP polytope.
Nice recent survey by Conforti-Cornuéjols-Zambelli (2010).

Our goal: to understand this phenomenon for convex optimization, not just LP.

"Extended formulations" in SDP

Many convex sets and functions can be modeled by SDP or SOCP in nontrivial ways. Among others:

- Sums of eigenvalues of symmetric matrices
- Convex envelope of univariate polynomials
- Multivariate polynomials that are sums of squares
- Unit ball of matrix operator and nuclear norms
- Geometric and harmonic means

E.g., Nesterov/Nemirovski, Boyd/Vandenberghe, Ben-Tal/Nemirovski, etc.

Often, clever and non-obvious reformulations.

“Extended formulations” in SDP

Many convex sets and functions can be modeled by SDP or SOCP in nontrivial ways. Among others:

- Sums of eigenvalues of symmetric matrices
- Convex envelope of univariate polynomials
- Multivariate polynomials that are sums of squares
- Unit ball of matrix operator and nuclear norms
- Geometric and harmonic means

E.g., Nesterov/Nemirovski, Boyd/Vandenberghe, Ben-Tal/Nemirovski, etc.

Often, clever and non-obvious reformulations.

Our questions

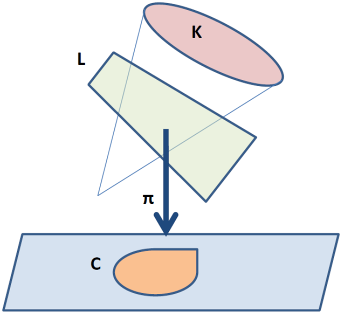
Existence and efficiency:

- When is a convex set representable by conic optimization?
- How to quantify the number of additional variables that are needed?

Given a convex set C, is it possible to represent it as

$$C = \pi(K \cap L)$$

where K is a cone, L is an affine subspace, and π is a linear map?



Cone lifts of convex bodies

When do such representations exist?

Even ignoring complexity aspects, this question is not well understood.

- Why a sphere is not a polytope?
- Can every basic closed semialgebraic set be represented using semidefinite programming?

What are “obstructions” for cone representability?

This talk: polytopes

What happens in the case of polytopes?

$$P = \{x \in \mathbb{R}^n \,:\, f_i^T x \leq 1\}$$

(WLOG, compact with $0 \in \operatorname{int} P$).

Polytopes have a finite number of facets f_i and vertices v_j.
Define a nonnegative matrix, called the *slack matrix* of the polytope:

$$[S_P]_{ij} = f_i^T v_j, \qquad i = 1, \ldots, |F| \quad j = 1, \ldots, |V|$$

This talk: polytopes

What happens in the case of polytopes?

$$P = \{x \in \mathbb{R}^n \,:\, f_i^T x \leq 1\}$$

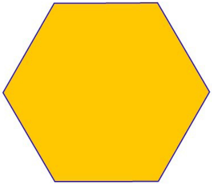
(WLOG, compact with $0 \in \operatorname{int} P$).

Polytopes have a finite number of facets f_i and vertices v_j.
Define a nonnegative matrix, called the *slack matrix* of the polytope:

$$[S_P]_{ij} = f_i^T v_j, \qquad i = 1, \ldots, |F| \quad j = 1, \ldots, |V|$$

Example: hexagon (I)

Consider a regular hexagon in the plane.



It has 6 vertices, and 6 facets. Its slack matrix is

$$S_H = \begin{pmatrix} 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

“Trivial” representation requires 6 facets. Can we do better?

Cone factorizations and representability

“Geometric” LP formulations exactly correspond to “algebraic” factorizations of the slack matrix.

For polytopes, this amounts to a *nonnegative factorization* of the slack matrix:

$$S_{ij} = \langle a_i, b_j \rangle, \qquad i = 1, \dots, v, \qquad j = 1, \dots, f$$

and a_i, b_i are nonnegative vectors.

Yannakakis (1991) showed that the minimal lifting dimension is equal to the nonnegative rank of the slack matrix.

Cone factorizations and representability

"Geometric" LP formulations exactly correspond to "algebraic" factorizations of the slack matrix.

For polytopes, this amounts to a *nonnegative factorization* of the slack matrix:

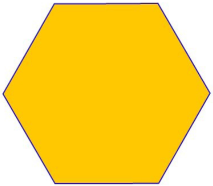
$$S_{ij} = \langle a_i, b_j \rangle, \qquad i = 1, \ldots, v, \qquad j = 1, \ldots, f$$

and a_i, b_i are nonnegative vectors.

Yannakakis (1991) showed that the minimal lifting dimension is equal to the nonnegative rank of the slack matrix.

Example: hexagon (II)

Regular hexagon in the plane.



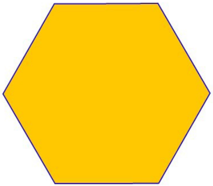
Slack matrix is

$$S_H = \begin{pmatrix} 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

Nonnegative rank is 5.

Example: hexagon (II)

Regular hexagon in the plane.



Slack matrix is

$$S_H = \begin{pmatrix} 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

Nonnegative rank is 5.

Bounding nonnegative rank

Want techniques to *lower bound* the nonnegative rank of a matrix.

In applications, these bounds may yield:

- Minimal size of latent variables
- Complexity lower bounds on extended representations

Known bounds exist (e.g. rank bound, combinatorial bounds, etc.).
Want to do better, using convex optimization...

Two convex cones

Two important and well-known convex cones of symmetric matrices:

- Copositive matrices:

$$\mathcal{C} := \{M \in \mathcal{S}^n : x^T M x \geq 0, \quad \forall x \geq 0\}$$

- Completely positive matrices:

$$\mathcal{B} := \mathrm{conv}\{xx^T : x \geq 0\}$$

These are proper cones (convex, closed, proper and solid), and they are dual to each other:

$$\mathcal{C}^* = \mathcal{B}, \qquad \mathcal{B}^* = \mathcal{C}.$$

Two convex cones

Two important and well-known convex cones of symmetric matrices:

- Copositive matrices:

$$\mathcal{C} := \{M \in \mathcal{S}^n : x^T M x \geq 0, \quad \forall x \geq 0\}$$

- Completely positive matrices:

$$\mathcal{B} := \operatorname{conv}\{xx^T : x \geq 0\}$$

These are proper cones (convex, closed, proper and solid), and they are dual to each other:

$$\mathcal{C}^* = \mathcal{B}, \qquad \mathcal{B}^* = \mathcal{C}.$$

A convex bound for nonnegative rank

Let $A \in \mathbb{R}_{+}^{m\times n}$ be a nonnegative matrix, and define

$$\nu_{+}(A) := \max_{W\in\mathbb{R}^{m\times n}} \left\{ \langle A, W\rangle \;:\; \begin{bmatrix} I & -W \\ -W^T & I \end{bmatrix} \text{ copositive } \right\}.$$

Then,

$$\mathrm{rank}_{+}(A) \geq \left(\frac{\nu_{+}(A)}{\|A\|_F}\right)^2,$$

where $\|A\|_F := \sqrt{\sum_{i,j} A_{i,j}^2}$ is the Frobenius norm of A.

- Essentially, a kind of “nonnegative nuclear norm”
- Convex, but hard... (membership in $\mathcal{B}$ and $\mathcal{C}$ is NP-hard!)

But, we know how to approximate them...

Proof

If $A = \sum_{i=1}^r u_i v_i^T$, a scaling argument show that wlog we can take $\|u_i\| = \|v_i\|$ for all i. By Cauchy-Schwarz,

$$\frac{\sum_{i=1}^r \|u_i\|\|v_i\|}{\sqrt{\sum_{i=1}^r \|u_i\|^2\|v_i\|^2}} \leq \sqrt{r} = \sqrt{\text{rank}_+(A)}$$

We can then bound the numerator and denominator:

- Numerator: if W is feasible, then $u_i^T W v_i \leq \|u_i\|\|v_i\|$, and thus $\langle A, W\rangle \leq \sum_{i=1}^r \|u_i\|\|v_i\|$.
- Denominator:

$$\|A\|_F^2 = \sum_{i,j=1}^r \langle u_i v_i^T, u_j v_j^T \rangle \geq \sum_{i=1}^r \|u_i\|^2\|v_i\|^2.$$

Approximation

Can approximate the cones $\mathcal{C}$ and $\mathcal{B}$ using sum of squares and semidefinite programming (P. 2000). We can write $\mathcal{C}$ as

$$\mathcal{C} = \left\{ M \in \mathcal{S}^n \; : \; \text{the polynomial } \sum_{i,j=1}^{n} M_{i,j} x_i^2 x_j^2 \text{ is nonnegative} \right\}.$$

The kth order relaxation is defined as:

$$\mathcal{C}^{[k]} = \left\{ M \in \mathcal{S}^n \; : \; \left(\sum_{i=1}^{n} x_i^2 \right)^k \left(\sum_{i,j=1}^{n} M_{i,j} x_i^2 x_j^2 \right) \text{ is a sums-of-squares} \right\}.$$

Clearly, $\mathcal{C}^{[k]} \subseteq \mathcal{C}$ and also $\mathcal{C}^{[k]} \subseteq \mathcal{C}^{[k+1]}$. Furthermore, each $\mathcal{C}^{[k]}$ is computable via SDP.

Simplest case ($k = 0$)

The case $k = 0$ is the simple sufficient condition for copositivity

$$M = P + N, \qquad P \succeq 0, \quad N_{ij} \geq 0.$$

Thus, the quantity $\nu_+^{[0]}(A)$ takes the more explicit form:

$$\nu_+^{[0]}(A) = \max \left\{ \langle A, W \rangle \ : \ \begin{bmatrix} I & -W \\ -W^T & I \end{bmatrix} \in \mathcal{N}^{n+m} + \mathcal{S}_+^{n+m} \right\}$$

For any $k \geq 0$:

$$\nu(A) \leq \nu_+^{[0]}(A) \leq \nu_+^{[k]}(A) \leq \nu_+(A) \leq \sqrt{\mathrm{rank}_+(A)} \|A\|_F$$

where $\nu(A)$ is the standard nuclear norm.

Simplest case ($k = 0$)

The case $k = 0$ is the simple sufficient condition for copositivity

$$M = P + N, \qquad P \succeq 0, \quad N_{ij} \geq 0.$$

Thus, the quantity $\nu_+^{[0]}(A)$ takes the more explicit form:

$$\nu_+^{[0]}(A) = \max \left\{ \langle A, W \rangle \ : \ \begin{bmatrix} I & -W \\ -W^T & I \end{bmatrix} \in \mathcal{N}^{n+m} + \mathcal{S}_+^{n+m} \right\}$$

For any $k \geq 0$:

$$\nu(A) \leq \nu_+^{[0]}(A) \leq \nu_+^{[k]}(A) \leq \nu_+(A) \leq \sqrt{\text{rank}_+(A)} \|A\|_F$$

where $\nu(A)$ is the standard nuclear norm.

Comparison: Rank bound

Trivially, $\text{rank}(A) \leq \text{rank}_+(A)$. Can our bound improve on this?

Consider

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

It is known that $\text{rank}(A) = 3$ and $\text{rank}_+(A) = 4$.

We have $\nu_+^{[0]}(A) = 4\sqrt{2}$, and thus our lower bound is sharp:

$$4 = \text{rank}_+(A) \geq \left(\frac{\nu_+^{[0]}(A)}{\|A\|_F} \right)^2 = \left(\frac{4\sqrt{2}}{\sqrt{8}} \right)^2 = 4.$$

Comparison: Boolean rank (rectangle covering)

A lower bound used in communication complexity. Relies only on sparsity pattern of matrix.

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

The rectangle covering number of A is 2 since supp(A) can be covered with the two rectangles $\{1,2\} \times \{2,3,4\}$ and $\{2,3,4\} \times \{1,2\}$.

Our bound yields $\text{rank}_+(A) \geq \lceil (\nu_+^{[0]}(A)/\|A\|_F)^2 \rceil = 3$. In fact $\text{rank}_+(A)$ is exactly equal to 3:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Example: hypercube

What is the extension complexity of the n-dimensional hypercube?
Is there better representation than the “obvious” $2n$ inequalities?

Rank bound is $n+1$. Goemans’ face-counting lower bound gives $\approx \sqrt{3}n$... Perhaps something nontrivial can be done?

Notice that the slack matrix is exponentially large ($2n \times 2^n$).

Proposition: Let $C_n = [0,1]^n$ be the hypercube in n dimensions and let $S(C_n) \in \mathbb{R}^{2n \times 2^n}$ be its slack matrix. Then

$$\mathrm{rank}_+(S(C_n)) = \left(\frac{\nu_+^{[0]}(S(C_n))}{\|S(C_n)\|_F}\right)^2 = 2n.$$

Example: hypercube

What is the extension complexity of the n-dimensional hypercube?
Is there better representation than the "obvious" $2n$ inequalities?

Rank bound is $n+1$. Goemans' face-counting lower bound gives $\approx \sqrt{3}n$... Perhaps something nontrivial can be done?

Notice that the slack matrix is exponentially large ($2n \times 2^n$).

Proposition: Let $C_n = [0,1]^n$ be the hypercube in n dimensions and let $S(C_n) \in \mathbb{R}^{2n\times 2^n}$ be its slack matrix. Then

$$\operatorname{rank}_+(S(C_n)) = \left(\frac{\nu_+^{[0]}(S(C_n))}{\|S(C_n)\|_F}\right)^2 = 2n.$$

Beyond LPs and nonnegative factorizations

LPs are nice, but what about broader representability questions?

In [GPT11], a generalization of Yannakakis' theorem to the general convex case. General theme:

"Geometric" extended formulations exactly correspond to "algebraic" factorizations of a slack operator.

polytopes/LP	convex sets/convex cones
slack matrix	slack operators
facets, vertices	primal and dual extreme points
nonnegative factorizations	conic factorizations

Polytopes and PSD factorizations

Even for polytopes, PSD factorizations can be interesting.

Well-known example: the *stable set* or *independent set* polytope.
Efficient SDP representations, but no known subexponential LP.

Natural notion: *positive semidefinite rank* ([GPT 11]).
Exactly captures the complexity of SDP-representability.

Polytopes and PSD factorizations

Even for polytopes, PSD factorizations can be interesting.

Well-known example: the *stable set* or *independent set* polytope.
Efficient SDP representations, but no known subexponential LP.

Natural notion: *positive semidefinite rank* ([GPT 11]).
Exactly captures the complexity of SDP-representability.

PSD rank of a nonnegative matrix

Let $M \in \mathbb{R}^{m \times n}$ be a nonnegative matrix.

The *PSD rank* of M, denoted rank_{psd}, is the smallest r for which there exists $r \times r$ PSD matrices $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_n\}$ such that

$$M_{ij} = \text{trace}\, A_i B_j, \qquad i = 1, \ldots, m \quad j = 1, \ldots, n.$$

Natural generalization of nonnegative rank.

The PSD rank determines the "best" semidefinite lifting.

PSD rank of a nonnegative matrix

Let $M \in \mathbb{R}^{m \times n}$ be a nonnegative matrix.

The *PSD rank* of M, denoted rank_{psd}, is the smallest r for which there exists $r \times r$ PSD matrices $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_n\}$ such that

$$M_{ij} = \text{trace}\, A_i B_j, \qquad i = 1, \ldots, m \quad j = 1, \ldots, n.$$

Natural generalization of nonnegative rank.

The PSD rank determines the "best" semidefinite lifting.

Lower bounding PSD rank?

Currently extending our bound to PSD rank, since combinatorial methods (based on sparsity patterns) cannot possibly work.

But, a few unexpected difficulties...

- In the PSD case, the underlying norm is non-atomic, and the corresponding “obvious” inequalities do not hold...
- “Noncommutative” versions of $\mathcal{C}$ and $\mathcal{B}$, quite complicated structure...

Nice links between rank_{psd} and quantum communication complexity, mirroring the situation between rank_+ and classical communication complexity (e.g., Fiorini *et al.* (2011), Jain *et al.* (2011), Zhang (2012)).

Computation

Even for nonnegative factorization, non-convex and very difficult.

A simple approach: alternating convex minimization.

For instance, for PSD factorizations of a nonnegative matrix $M = AB$, we can alternate between minimizing over $A = [A_1, \ldots, A_m]^T$ and $B = [B_1, \ldots, B_n]$:

$$\underset{A_i \succeq 0}{\text{minimize}} \, \|M - AB\| \qquad \underset{B_i \succeq 0}{\text{minimize}} \, \|M - AB\|$$

These subproblems are SDPs (and if $\|\cdot\|$ is the Euclidean norm, they are decoupled). However, no global guarantees.

Ongoing work of F. Glineur (UCL).

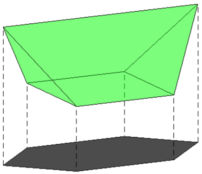
The End

Thank You!

Want to know more?

- J. Gouveia, P.A. Parrilo, R. Thomas, Lifts of convex sets and cone factorizations, *Mathematics of Operations Research*, to appear, 2013. `arXiv:1111.3164`.
- H. Fawzi, P.A. Parrilo, New lower bounds on nonnegative rank using conic programming, `arXiv:1210.6970`.

Example: hexagon (III)



A nonnegative factorization:

$$S_H = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 & 2 & 1 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$