Mathematics for Economists: A Synopsis

Juuso Välimäki

Spring 2020

This synopsis outlines without any explanations what the main results in the course are.

1 Linear models

For $x \in \mathbb{R}$, and $a, b \in \mathbb{R}$:

$$ax = b.$$

If $a = 0$ and $b \neq 0$, no solution. If $a \neq 0$ and $b = 0$, a single solution. Infinitely many solutions if $a = b = 0$.

For $x \in \mathbb{R}^n$, and A an $m \times n$ matrix, $b \in \mathbb{R}^m$,

$$Ax = b.$$

A solution exists is A has rank at least m. The solution is unique if $n = m$. If rank less than m, then solutions do not exist for some b.

Most important case: $n = m$. Then solution exists and is unique for all b if and only if A has rank A. In this case $\det(A) \neq 0$ and A^{-1} exists. In this case:

$$x = A^{-1}b.$$

Gaussian elimination and determinants are the main methods for determining whether A has full rank. Gaussian elimination also gives the x

solving the equation. Cramer's rule is an alternative way for computing the solution:

$$x_i = \frac{\det(A_i(b))}{\det(A)},$$

where $A_i(b)$ is the matrix obtained by replacing the i th column of A by b.

2 The derivative

Consider $f : \mathbb{R}^n \to \mathbb{R}$. Consider an arbitrary point $\hat{x} \in \mathbb{R}^n$. Fix all the other coordinates, but let x_i vary. Then we have the function

$$f(\hat{x}_1, ...\hat{x}_{i-1}, x_i, \hat{x}_{i+1}, ..., x_n)$$

of the single real variable x_i. The derivative of this function at $x_i = \hat{x}_i$ is called the *partial derivative of f with respect to x_i at $\hat{x}$* and denoted by

$$\frac{\partial f(\hat{x})}{\partial x_i}.$$

The derivative of f at $\hat{x}$ is a linear function that approximates f well for x close to $\hat{x}$. If partial derivatives with respect to all x_j exist at $\hat{x}$ and if the partial derivatives are continuous at $\hat{x}$, then f has a derivative at $\hat{x}$. In this case, the derivative is the row vector

$$D_x f(\hat{x}) = (\frac{\partial f(\hat{x})}{\partial x_1}, ... \frac{\partial f(\hat{x})}{\partial x_n}).$$

We have:

$$f(x) - f(\hat{x}) = D_x f(\hat{x})(x - \hat{x}) + h.o.t.,$$

where h.o.t. means terms that vanish in comparison to the term $D_x f(\hat{x})(x - \hat{x})$ as x is close to $\hat{x}$.

For small Δx, we can compute the approximation:

$$f(\hat{x} + \Delta x) - f(\hat{x}) \approx D_x f(\hat{x}) \Delta x.$$

We call this the directional derivative of f in directions Δx at $\hat{x}$.

The gradient $\nabla f(\hat{x})$ is the transpose of the derivative (i.e. the column vector of partial derivatives at $\hat{x}$.

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is a vector of functions $f_i : \mathbb{R}^n \to \mathbb{R}$:

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

If all the partial derivatives of all component functions exist and are continuous at $\hat{x}$, then the derivative of such a function at $\hat{x}$ is the $m \times n$ matrix:

$$D_x f(x) = \begin{pmatrix} \frac{\partial f_1(\hat{x})}{\partial x_1} & \dots & \frac{\partial f_1(\hat{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\hat{x})}{\partial x_1} & \dots & \frac{\partial f_m(\hat{x})}{\partial x_n} \end{pmatrix}.$$

3 Implicit function theorem

Consider functions $f : R^{n+m} \to \mathbb{R}^n$. Let $y \in \mathbb{R}^n$ denote endogenous variables in an economic model and $x \in \mathbb{R}^m$ the exogenous variables and write the model as:

$$f(y, x) = 0.$$

Assume that $(\hat{y}, \hat{x})$ satisfies the model, i.e. $(f(\hat{y}, \hat{x}) = 0$. We want to know how the endogenous variables y behave when x changes a bit from $\hat{x}$.

Assume that the derivative of f exists at $(\hat{y}, \hat{x})$ and that the derivative with respect to endogenous variables

$$D_y f(\hat{y}, \hat{x}) = \begin{pmatrix} \frac{\partial f_1(\hat{y},\hat{x})}{\partial y_1} & \dots & \frac{\partial f_1(\hat{y},\hat{x})}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\hat{y},\hat{x})}{\partial y_1} & \dots & \frac{\partial f_m(\hat{y},\hat{x})}{\partial y_n} \end{pmatrix}$$

has full rank (i.e. non-zero determinant). Then the implicit function theorem tells us that we have a differentiable implicit function $y(x)$ with $y(\hat{x}) = \hat{y}$ defined in a neighborhood of $\hat{x}$ such that for all x in that neighborhood,

$$f(y(x), x) = 0.$$

Chain rule gives:

$$D_y f(\hat{y},\hat{x}) D_x y(x) + D_x f(\hat{y},\hat{x}) = 0.$$

Since we have assumed that D_y has full rank, it is invertible and:

$$D_x y(\hat{x}) = -(D_y f(\hat{y},\hat{x}))^{-1} D_x f(\hat{y},\hat{x}).$$

The effect on y_i of a change in a particular x_k can be computed with Cramer's rule.

Illustration: indifference curves

On an indifference curve, $u(x,y) = \overline{u}$. What is the slope of the indifference curve at point $(\hat{x},\hat{y})$. The level of utility at that point is $u(\hat{x},\hat{y})$. On the same indifference curve, the level of utility is the same:

$$u(x,y) = u(\hat{x},\hat{y}).$$

If $\frac{\partial u(\hat{x},\hat{y})}{\partial y} \neq 0$, then implicit function theorem tells us that for x near $\hat{x}$, we have a function $y(x)$ with $y(\hat{x}) = \hat{y}$ and

$$y'(\hat{x}) = -\frac{\frac{\partial u(\hat{x},\hat{y})}{\partial x}}{\frac{\partial u(\hat{x},\hat{y})}{\partial y}}.$$

4 Unconstrained optimization

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. A point $\hat{x}$ is a *maximum* of f if $f(\hat{x}) \geq f(x)$ for all x. It is a *minimum* if $f(\hat{x}) \leq f(x)$ for all x.

The point $\hat{x}$ is called a *local maximum* (*local minimum*) if $f(\hat{x}) \geq (\leq) f(x)$ for all x in some neighborhood of $\hat{x}$.

A necessary condition for local (and therefore also global) minimum and maximum is that all the partial derivatives of f vanish at $\hat{x}$ or $\nabla f(\hat{x}) = 0$. Necessary condition means that all minima and maxima have this property.

Points satisfying the necessary condition are called critical points of f. In order to classify the critical points, we need to look at the Hessian matrix of second derivatives of f at $\hat{x}$.

Second order Taylor approximation gives:

$$f(x) - f(\hat{x}) = D_x f(\hat{x})(x - \hat{x}) + \frac{1}{2}(x - \hat{x}) \cdot H_x f(\hat{x})(x - \hat{x}) + h.o.t.,$$

where

$$H_x f(\hat{x}) = \begin{pmatrix} \frac{\partial^2 f(\hat{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\hat{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\hat{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\hat{x})}{\partial x_n^2} \end{pmatrix},$$

and the h.o.t. vanish in comparison to the first and second order terms for x close to $\hat{x}$.

At a critical point, $D_x f(\hat{x}) = 0$ so that the first term on the right-hand side vanishes. By Young's theorem, the Hessian matrix is a symmetric matrix and hence defines a quadratic form.

A necessary condition for a local maximum (minimum) at a critical point $\hat{x}$ is that for all x,

$$(x - \hat{x}) \cdot H_x f(\hat{x})(x - \hat{x}) \leq (geq)0.$$

A sufficient condition for local maximum or at $\hat{x}$ is that the inequality above is strict for all x.

A symmetric matrix is said to be positive (negative) definite if $x \cdot Ax > (<)0$ for all x. It is positive (negative) semi-definite if the $x \cdot Ax \geq (\leq)0$ for all x.

If $x \cdot Ax$ is strictly positive for some x and strictly negative for some x, A is said to be indefinite. The criteria for checking definiteness are given in the notes.

Notice that for functions of a single real variable, the Hessian is just the second derivative of the function.

5 Convex and concave functions

A set X is concave if the line connecting any two points x, y in the set also belongs to X. A concave function is a function on a convex set, whose graph along the line connecting x, y lies above the line joining $f(x)$ and $f(y)$, i.e. for all $x, y \in X$ and all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y).$$

A function is convex if for all $x, y \in X$ and $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Every critical point of a concave (convex) function f is a global maximum (minimum) of f.

A twice differentiable function f is concave (convex) if and only if the Hessian at x is negative (positive) semi-definite for all x.

If $g(x,a)$ are linear functions of a, then $f(a) = \max_x g(x,a)$ is a convex function of a. Similarly, the minimum of linear functions is concave.

A function f on a convex set X is quasiconcave if for all $x, y \in X$ and all $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}.$$

It is quasiconvex if

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}.$$

A function f is quasiconcave (quasiconvex) if and only if its upper (lower) contour set is convex.

Any critical point of a quasiconcave (quasiconvex) function with non-vanishing derivative is a global maximum (minimum).

6 Optimization with equality constraints

Consider the maximization problem:

$$\max f(x_1, ..., x_n)$$

subject to

$$\begin{aligned} g_1(x_1, ..., x_n) &= 0, \\ &\vdots \\ g_k(x_1, ..., x_n) &= 0. \end{aligned}$$

If f is continuous and the feasible set is compact, then Weierstrass' theorem quarantees that a maximum exists.

Constraint qualification at x requires that the matrix of the derivatives $Dg(x)$ has full rank at x. For a single constraint, this just requires a non-zero gradient.

With a single constraint, consider the following intuition. The feasible set cannot intersect the level curve of the objective function at optimum. If there was such an intersection, part of the feasible set would give a strictly higher value to the objective function than the intersection point. Therefore an intersection point cannot be optimum and we get the tangency of the level curve and the feasible set. Since the feasible set is the level curve of the constraint function, the gradients of the objective function and the constraint must be collinear.

A necessary condition for a maximum at a point $\hat{x}$ where constraint qualification holds is that $\hat{x}, \hat{\mu}$ be a critical point of the following Lagrangean function

$$\mathcal{L}(x, \mu) = f(x) - \sum_{j=1}^{k} \mu_j g_j(x).$$

In other words, we must look for $(\hat{x}, \hat{\mu})$ such that

$$\frac{\partial f(\hat{x})}{\partial x_i} - \sum_{j=1}^{k} \hat{\mu}_j \frac{\partial g_j(\hat{x})}{\partial x_i} = 0 \text{ for all } i \in \{1, ..., n\},$$

$$g_j(x_1, ..., x_n) = 0 \text{ for all } j \in \{1, ..., k\}.$$

For a linear constraint, the feasible set is convex and therefore by the results in the previous section, such critical points are global maxima.

Otherwise, one must check whether the critical points are maxima or minima using the bordered Hessian at the critical point.

7 Optimization with inequality constraints

The problem is now to

$$\max f(x_1, ..., x_n)$$

subject to

$$g_1(x_1, ..., x_n) \leq 0,$$
$$\vdots$$
$$g_k(x_1, ..., x_n) \leq 0.$$

The difficulty is now that we do not know which of the constraints are binding. The necessary conditions can be formulated using a Lagrangean function, but now we need also complementary slackness conditions for the constraints. If the constraint binds, its multiplier is positive, if it does not bind, the multiplier is negative. For inequality constraints, we also get the non-negativity of the multipliers as a necessary condition.

The Lagrangean of the problem is:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{j=1}^{k} \lambda_j g_j(x).$$

The necessary first-order Kuhn-Tucker conditions for the problem at a point where the constraint qualification holds (i.e. the derivatives of the binding constraints are linearly independent) are given by:

$$\frac{\partial f(\hat{x})}{\partial x_i} - \sum_{j=1}^{k} \hat{\mu}_j \frac{\partial g_j(\hat{x})}{\partial x_i} = 0 \text{ for all } i \in \{1, ..., n\},$$
$$\lambda_j g_j(x) = 0 \text{ for all } j \in \{1, ..., k\},$$
$$g_j(x_1, ..., x_n) \leq 0 \text{ for all } j \in \{1, ..., k\},$$
$$\lambda_j \geq 0 \text{ for all } j \in \{1, ..., k\}.$$

If the objective function is quasiconcave and has a nonzero gradient and if the feasible set is convex, then any point satisfying the K-T conditions is a global maximum.

Of course, it is a nightmare to solve a system such as this one in general. Therefore it is advisable to look carefully at the problem and argue which of the constraints cannot bind at optimum (this can often be done for non-negativity constraints) and argue which constraints are binding (typically budget constraints). The art of solving optimization problems is often the art of simplifying the constraints.

8 Utility maximization: an example

Consider the problem of maximizing

$$u(x,y) = (a_x x^\rho + a_y y^\rho)^{\frac{1}{\rho}},$$

subject to

$$p_x x + p_y y \le w, x, y \ge 0,$$

where $\rho < 1, \quad \rho \neq 0$ and $a_x, a_y > 0$.

The constraint set is compact and the utility function is convex so therefore an optimum exists by Weierstrass' theorem. The utility function is quasiconcave with non-vanishing gradient and the feasible set is convex and thus K-T conditions are also sufficient. The binding constraints are also linearly independent (since the non-negativity constraints are orthogonal and the budget constraint cannot bind simultaneously with all the other constraints).

Lagrangean for the problem:

$$\mathcal{L}(x,y,\lambda,\lambda_x,\lambda_y) = (a_x x^\rho + a_y y^\rho)^{\frac{1}{\rho}} - \lambda(p_x x + p_y y - w) + \lambda_x x + \lambda_y y.$$

Kuhn-Tucker conditions:

$$\frac{\partial u(x,y)}{\partial x} - \lambda p_x + \lambda_x = \rho a_x x^{\rho-1}\frac{1}{\rho}(a_x x^\rho + a_y y^\rho)^{\frac{1}{\rho}-1} - p_x + \lambda_x = 0,$$

$$\frac{\partial u(x,y)}{\partial y} - \lambda p_y + \lambda_y = \rho a_y y^{\rho-1}\frac{1}{\rho}(a_x x^\rho + a_y y^\rho)^{\frac{1}{\rho}-1} - \lambda p_y + \lambda_y = 0,$$

$$\lambda(p_x x + p_y y - w) = 0,$$

$$\lambda_x x = \lambda_y y = 0,$$

$$\lambda, \lambda_x, \lambda_y, x, y \ge 0.$$

Since $\rho < 1$, $\frac{\partial u(0,y)}{\partial x}$ and $\frac{\partial u(x,0)}{\partial y}$ are not defined and we see that the first and second lines in K-T conditions cannot be satisfied at $x = 0$ or $y = 0$.

Therefore we know that $\lambda_x = \lambda_y = 0$. Budget constraint must bind since $\frac{\partial u(x,y)}{\partial x}, \frac{\partial u(x,y)}{\partial y} > 0$ for all $(x, y) > 0$.

Solving for λ from the first two lines, setting equal and cross-multiplying gives:

$$\frac{\frac{\partial u(x,y)}{\partial x}}{\frac{\partial u(x,y)}{\partial y}} = \frac{p_x}{p_y}.$$

Also, the budget constraint holds with equality:

$$p_x x + p_y y = w.$$

Plugging in the marginal utilities gives:

$$\frac{a_x x^{\rho-1}}{a_y y^{\rho-1}} = \frac{p_x}{p_y},$$

or

$$\frac{x}{y} = (\frac{a_y p_x}{a_x p_y})^{\frac{1}{\rho-1}},$$

or

$$y = x(\frac{a_y p_x}{a_x p_y})^{\frac{1}{1-\rho}}. \tag{1}$$

Substituting into the budget constraint, we get:

$$p_x x + p_y x(\frac{a_y p_x}{a_x p_y})^{\frac{1}{1-\rho}} = w.$$

We can solve for x_1 to get

$$x = \frac{w}{p_x + p_y(\frac{a_y p_x}{a_x p_y})^{\frac{1}{1-\rho}}}.$$

Substituting this into (1) lets us solve for y:

$$y = \frac{w}{p_y + p_x(\frac{a_x p_y}{a_y p_x})^{\frac{1}{1-\rho}}}.$$

9 Value function and envelope theorem

Consider an unconstrained maximization problem of a function of a single real variable x, where the objective function depends on a parameter $\alpha \in \mathbb{R}$.

$$\max_{x\in\mathbb{R}} f(x,\alpha).$$

Let $x(\alpha)$ be the solution to this problem. Consider the maximum value of the objective function that is achievable at the parameter $\hat{\alpha}$

$$V(\alpha) := f(x(\alpha),\alpha).$$

At the (unconstrained) optimum $x(\hat{\alpha})$, by the first-order condition:

$$\frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial x} = 0.$$

Compute the change in V from a change in the parameter:

$$V'(\hat{\alpha}) = \frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial x} x'(\hat{\alpha}) + \frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial \alpha},$$

since $\frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial x} = 0$.

This observation is called the envelope theorem. For $x \in \mathbb{R}^n$, the message is exactly the same. The first order-condition is now:

$$\frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial x_i} = 0 \text{ for all } i \in \{1,...,n\}.$$

Assuming the conditions for implicit function theorem, we have by chain rule:

$$V'(\hat{\alpha}) = \sum_{i=1}^{n} \frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial x_i} x_i'(\hat{\alpha}) + \frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial \alpha}.$$

Again, the first term vanishes by first-order condition and we are left with

$$V'(\hat{\alpha}) = \frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial \alpha}.$$

Suppose that we have an equality constrained parametric maximization problem for $x \in \mathbb{R}^n$:

$$\max_x f(x, \alpha)$$
$$\text{subject to } g(x, \alpha) = 0.$$

The value function is still defined as before:

$$V(\alpha) = f(x(\alpha), \alpha).$$

Begin the analysis by forming the Lagrangean:

$$\mathcal{L}(x, \mu; \alpha) = f(x, \alpha) - \mu g(x, \alpha).$$

The envelope theorem relates the derivative of the value function with respect to the parameter to the partial derivatives of the Lagrangean.

Theorem 1 (Envelope theorem)**.** In an optimization problem subject to an equality constraint, we have:

$$V'(\alpha) = \frac{\partial \mathcal{L}(x, \mu; \alpha)}{\partial \alpha}.$$

The envelope theorem gives us a nice way of understanding the Lagrange multipliers in utility maximization problems. The Lagrangean for the UMP with a single binding equality constraint is:

$$\mathcal{L}(x, \lambda) = u(x) - \mu \left[\sum_{i=1}^{n} p_i x_i - w \right].$$

The maximum value function

$$v(p, w) = \max u(x) \text{ subject to } p \cdot x = w,$$

is called the indirect utility function. It computes the optimal utility level for all combinations of prices $p \in \mathbb{R}^n_{++}$ and income $w > 0$.

Envelope theorem tells us that:

$$\frac{\partial v(p, w)}{\partial w} = \mu.$$

The multiplier tells that if income goes up (down) by dw, then utility goes up (down) by λdw. Because of this the multiplier is called the shadow price of income.

10 Difference equations

Difference equations relate the future state x_{n+1} of a dynamical system to its current state x_n:

$$x_{n+1} = f(x_n) \text{ for all } t \in \mathbb{Z}_+.$$

The function f is the system equation. The unknowns in the problem is the sequence $\{x_n\}_{n=0}^{\infty}$ satisfying the system equation.

The simplest form of difference equations are linear difference equations with constant coefficients. These can be written as:

$$x_{n+1} = Ax_n + b_n,$$

where b_n is a given sequence. If $b_n = 0$ for all n, we have a homogenous equation. We start with the simplest homogenous equations where $x_n \in \mathbb{R}$ and $A = a \in \mathbb{R}$.

Solving the homogenous equation is very easy. If $x_{n+1} = ax_n$ for all n, then $x_{n+k} = a^k x_n$. Hence any sequence of the form $x_n = ca^n$ solves the difference equation. If we are given the initial value x_0, the solution is $x_n = x_0 a^n$. In other words, the initial value pins down the coefficient c of the general solution.

Consider next an inhomogenous equation,

$$x_{n+1} = ax_n + b,$$

where $b_n = b$ for all n. Clearly the constant solution $x_n = \frac{b}{1-a}$ for all n solves the equation. I claim that also $x_n = ca^n + \frac{b}{1-a}$ solves the equation. But this follows immediately from the fact that $ca^{n+1} = aca^n$.

This principle holds more generally. If you have a particular solution x_n^P to the inhomogenous equation and the general solution of the homogenous equation x_n^H, then the general solution to the problem is $x_n^P + x_n^H$. This is called the principle of superposition and it arises from the linearity of the equations in x_{n+1}, x_n. It is valid also for the case with $x_n \in \mathbb{R}^n$.

Consider next linear systems with constant coefficients. Let $x_n \in \mathbb{R}^k$ for all n and let A be an $k \times k$ matrix of real numbers. A linear homogenous system is then given by:

$$x_{n+1} = Ax_n.$$

As before, we can 'solve' this by repeated substitution to get

$$x_{n+k} = A^k x_n.$$

Hence I could write the general solution as $x_n = A^n c$ for some vector $c = (c_1, ..., c_k)$. I do not consider this a real solution since it is almost impossible to see what A^n is except in some very special cases. If A is a diagonal matrix with diagonal elements $a_1, ..., a_k$, then the solution becomes

$$x_{i,n} = ca_i^n \text{ for } i \in \{1, ..., k\}.$$

Here we have essentially independent variables and the difference equation for each can be solved separately.

To deal with the general case, we want to change the basis in $\mathbb{R}^k$ so that A is diagonal in that basis. This involves the eigenvectors and eigenvalues of A. You can visualize the effect of matrix multiplication on vectors as consisting of two operations: a rotation and a stretching or shrinking. Eigenvectors of A are those vectors that are not rotated, i.e. if $x \neq 0$ is an eigenvector of A, then for some $\lambda \in \mathbb{R}$,

$$Ax = \lambda x.$$

We may write this more compactly as

$$(A - \lambda I)x = 0,$$

where I is the $n \times n$ identity matrix. But from basic linear algebra, we know that a homogenous linear equation can have a non-zero solution only if the matrix does not have full rank, i.e. if $\det(A - \lambda I) = 0$. The values of λ for which this determinant is zero are called the eigenvalues of A.

The determinant of $A - \lambda I)$ is called the characteristic polynomial of A so the eigenvalues are the roots of the characteristic polynomial. If A has n distinct eigenvelues $\lambda_1, ..., \lambda_n$, then it has also n linearly independent eigenvectors $v_1, ..., v_n$ so that

$$Av_i = \lambda_i x_i.$$

I can express any $x \in \mathbb{R}^n$ given in the usual coordinate system in the coordinate system of spanned by the eigenvectors by simple matrix multiplication. Let $P = [v_1 \quad v_2 \quad \ldots v_n$ be the matrix formed by the eigenvectors.

Then for any vector y expressed in the coordinate system of the eigenvectors, we can translate it to the standard system by $x = Py$. Similarly any x in the standard system is $y = P^{-1}x$ in the system of the eigenvectors.

$$y_{n+1} = P^{-1}x_{n+1} = P^{-1}Ax_n = P^{-1}APy_n.$$

Now we want to show that $P^{-1}AP = \Lambda$, where Λ is the diagonal matrix of eigenvalues. But this is the same claim (as can be seen by premultiplying by P) as:

$$AP = P\Lambda.$$

But this follows immediately from the fact that P consists of the eigenvectors of A.

Hence we have: $y_n = (y_{1,n}, ..., y_{k,n}) = (c_1\lambda_1^n, ..., c_k\lambda_k^n)$. Since $x_n = Py_n$, we have the general solution:

$$x_n = c_1\lambda_1^n v_1 + ... + c_k\lambda_k^n v_k.$$

Note that $A^k = P\Lambda P^{-1}$. Therefore we could have also concluded that

$$x_n = P\lambda^k P^{-1}x_0.$$

The two methods give the same results since $Pc = x_0$ or $c = P^{-1}x_0$.