Solutions: second stage of Israeli students competition, 2015.

1. Find such $x>0$, for which $\int_0^x \frac{dt}{t^{1+\ln t}} = \int_x^\infty \frac{dt}{t^{1+\ln t}}$.

Answer. $x=1$.

First solution. We shall start with general remarks on convergence. When $t\to 0$ we have $\ln t < -2$, so $\frac{1}{t^{1+\ln t}} < t$, so integral is well-defined at 0. As $t\to\infty$, $\ln t > 2$, so $\frac{1}{t^{\ln t}} < \frac{1}{t^2}$, so the integral converges at ∞.

Notice, that the integrated function is positive, so as x is increasing, the left hand side increasing, and right hand side is decreasing. So there can be only one answer. Since the integral is well-defined at both ends, the LHS is sufficiently small when x is close to zero, and RHS is sufficiently small when x is large, so by continuity an answer exists.

Consider a substitution $s=\frac{1}{t}$. Then $\ln t = -\ln s$, and $dt = -\frac{ds}{s^2}$, but we can skip the minus sign if we revert the endpoints of the integral (which is a logical thing to do, since the substitution reverses the order. So we get

$$\int_{1/x}^{\infty} s^{1-\ln s}\frac{ds}{s^2} = \int_0^{1/x} s^{1-\ln s}\frac{ds}{s^2}$$

$$\int_{1/x}^{\infty} s^{1-\ln s}\frac{ds}{s^2} = \int_0^{1/x} s^{1-\ln s}\frac{ds}{s^2}$$

$$\int_{1/x}^{\infty} s^{1-\ln s}\frac{ds}{s^2} = \int_0^{1/x} s^{1-\ln s}\frac{ds}{s^2}$$

$$\int_{1/x}^{\infty} \frac{ds}{s^{1+\ln s}} = \int_0^{1/x} \frac{ds}{s^{1+\ln s}}$$

So, if x is an answer then $\frac{1}{x}$ is also an answer. But the answer is unique, so $x=\frac{1}{x}$, hence $x=1$.

Second solution. We will apply the substitution $\ln t = y$, which means $t=e^y$. Then $dt=\frac{dy}{y}$, and the new condition is $\int_{-\infty}^{\ln x} \frac{dy}{\left(e^y\right)^y} = \int_{\ln y}^{\infty} \frac{dx}{\left(e^y\right)^y}$.

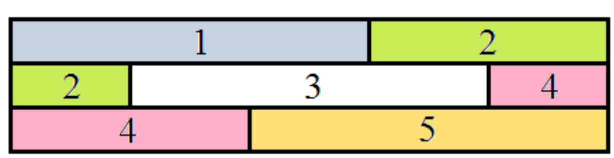
$$\int_{-\infty}^{\ln x} e^{-y^2} dy = \int_{\ln x}^{\infty} e^{-y^2} dy$$

The integral $\int e^{-y^2} dy$ is famous (especially in probability theory) , but it is not an elementary function. The function e^{-y^2} is even, positive and quickly decreasing, so it is obvious that the only point which cuts the integral in half is 0.
Hence $\ln x = 0$, and $x = 1$.

2. N people must travel from one end of the road to another. The length of the road is L. They have K bicycles ($K < N$). The velocity of walking man is v_1, and the velocity of a bicycle is v_2 (obviously, $v_1 < v_2$). How much time is required?

Answer. $\frac{L}{N}\left(\frac{K}{v_2} + \frac{N-K}{v_1}\right)$

Solution. We shall introduce natural coordinates on the road: the first end is zero, and the target end is L. The total displacement of the bicycles is KL at most, and that happens only if all bicycles make all the way from 0 to the target (if someone fancies riding a bicycle in the opposite direction for some reason, it is regarded as negative displacement). So, one of the people who was the least advanced by the bicycles, made $\frac{KL}{N}$ at most by the bicycle, and the rest of the way, $L - \frac{KL}{N}$ at least, by foot, so he spent no less than $\frac{KL}{N}\Big/ v_2 + \left(L - \frac{KL}{N}\right)\Big/ v_1 = \frac{L}{N}\left(\frac{K}{v_2} + \frac{N-K}{v_1}\right)$.
The hard part is to prove that this number can be achieved. It is easy to guess from the first part of the proof, that in order to transport all people in this amount of time, all of them must constantly move forward, and do precisely $\frac{K}{N}$ of the way by bicycle. We shall build a table of height K and length L, and we shall pack it with blocks of height 1 and length $\frac{KL}{N}$.



In the picture there is an example for $K = 3, N = 5$. The first block is in the first line but it doesn't take the whole line; each block starts in the same place, where the previous block stops, but if it is too long, then part of the block for which there is not enough place in the current line, is chopped away and moved to the beginning of the next line. So, all blocks have the same length, even if some are divided.

This table we've built is a schedule of bicycle usage (a term schedule usually means time-table, but in our case it is distance-table). Blocks correspond to people, lines of the table correspond to bicycles; the horizontal direction to the locations on the road. So, on this table we see which bicycle on which part of the road can be used by which person.

If someone arrives to a spot, where he (according to our schedule) has to take a bicycle, the previous owner of the bicycle enjoyed more bicycle-time than he, so he is more advanced along the road, so he has already left him a bicycle in precisely this spot. So this schedule can be implemented. Hence each person can do precisely $\frac{K}{N}$ of the way by bicycle, and then they all can arrive in the time we computed.

3. A unit cube in 4-dimensional Euclidean space contains a 3-dimensional Euclidean ball of radius R. What is the greatest possible value of R?

Answer. $\frac{1}{\sqrt{3}}$.

Solution. Coordinates in $\mathbb{R}^4$ will be denoted x_1, x_2, x_3, x_4, and we can assume the unit cube is $\left[-\frac{1}{2}, \frac{1}{2}\right]^4$. Consider a hyperplane $a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_4 \cdot x_4 = s$.

Without loss of generality, we can assume that $\sum_{i=1}^{4} a_i^2 = 1$. In short, we can describe the hyperplane by the equation $\langle n, x \rangle = s$, where $n = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$.

Consider also a vector $v = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ b_4 \end{pmatrix}$, such that $\langle v_4, n \rangle = 0$.

It means in coordinates that $0 = a_1^2 + a_2^2 + a_3^2 + a_4 b_4 = 1 - a_4^2 + a_4 b_4 = 0$, so $b_4 = \frac{a_4^2 - 1}{a_4} = a_4 - \frac{1}{a_4}$.

Therefore $|v|^2 = a_1^2 + a_2^2 + a_3^2 + b_4^2 = 1 - a_4^2 + a_4^2 - 2 + \frac{1}{a_4^2} = \frac{1}{a_4^2} - 1$.

We can find a three vectors v_1, v_2, v_3 which are of unit length, and orthogonal to each other and to n, such that $v_1 = \frac{v}{|v|}$. Since v_2, v_3 are orthogonal to both v and n, they are orthogonal both to $\frac{v+n}{2} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ 0 \end{pmatrix}$ and to $\frac{v-n}{2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ t \end{pmatrix}$, (where $t = -\frac{1}{a_4}$) so they have zero last coordinate. Any vector parallel to the hyperplane can be expressed as $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$, and the length of the vector is $\alpha_1^2 + \alpha_2^2 + \alpha_3^2$ but its projection to the x_4-axis is α_1 times the last coordinate of v_1. The diameter of the 3-dimensional ball of radius R in the hyperplane, which has the longest projection on the x_4-axis, is parallel to the vector $2Rv_1 = 2R\frac{v}{|v|}$, and its last coordinate is

$$2R\frac{b_4}{|v|} = {2R\left(a_4 - \frac{1}{a_4}\right)} \Big/ {\sqrt{\frac{1}{a_4^2} - 1}} = {2R\left(a_4^2 - 1\right)} \Big/ {\sqrt{1 - a_4^2}} = -2R\sqrt{1 - a_4^2}$$

To have the ball inside the unit cube, we should have $2R\sqrt{1 - a_4^2} \le 1$, therefore $\sqrt{1 - a_4^2} \le \frac{1}{2R}$, hence $1 - \frac{1}{4R^2} \le a_4^2$.

But similar argument holds for each coordinate, hence $1-\frac{1}{4R^2}\le a_i^2$, so

$4-\frac{1}{R^2}\le a_1^2+a_2^2+a_3^2+a_4^2=1$, hence $3\le\frac{1}{R^2}$, so $R\le\frac{1}{\sqrt{3}}$.

Just to be sure, let us verify, that a 3-dimensional ball of radius $\frac{1}{\sqrt{3}}$ can be inserted into the 4-dimensional unit cube. It is easy to guess the hyperplane; that is the case when all inequalities we wrote turn to the equalities. So, take the hyperplane $x_1+x_2+x_3+x_4=0$, and in it take a ball of radius $\frac{1}{\sqrt{3}}$. We have to verify that

$$\begin{cases} x_1+x_2+x_3+x_4=0 \\ x_1^2+x_2^2+x_3^2+x_4^2\le\frac{1}{3} \end{cases}$$

implies $-\frac{1}{2}\le x_i\le\frac{1}{2}$ for each i, but by symmetry it is enough to verify for $i=4$ by symmetry, also, it is possible to revert sign of all x_i simultaneously, and so it is enough to show that $x_4\le\frac{1}{4}$.

Obviously $(x_1-x_2)^2+(x_1-x_3)^2+(x_2-x_3)^2\ge 0$, hence

$$2\left(x_1^2+x_2^2+x_3^2\right)\ge 2\left(x_1x_2+x_1x_3+x_2x_3\right)$$

$$3\left(x_1^2+x_2^2+x_3^2\right)\ge x_1^2+x_2^2+x_3^2+2\left(x_1x_2+x_1x_3+x_2x_3\right)=\left(x_1+x_2+x_3\right)^2$$

$$x_1^2+x_2^2+x_3^2\ge\frac{1}{3}\cdot\left(x_1+x_2+x_3\right)^2=\frac{1}{3}\cdot\left(-x_4\right)^2=\frac{x_4^2}{3}$$

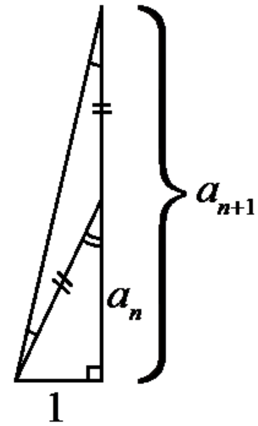
$$\frac{1}{3}\ge x_1^2+x_2^2+x_3^2+x_4^2\ge\frac{4x_4^2}{3}$$

$$\frac{1}{4}\ge x_4^2.$$

Q.E.D.

Remark. We could argue that in our example all inequalities in the first discussion turn into equalities, and skip some algebra, but it is good to verify an argument in independent way and so to make sure that we didn't have an arithmetic mistake.

4. The sequence $\{a_n\}$ is defined by recurrent formula $a_{n+1} = a_n + \sqrt{1+a_n^2}$, and $a_1 = 1$. Compute $\lim\limits_{n\to\infty} \frac{2^n}{a_n}$.



Answer. $\frac{2}{\pi}$.

First solution. The formula becomes clear, if you look at its geometric meaning. Construct a right-angled triangle, the short sides of which are 1 and a_n, and the long side is, by Pythagoras theorem, $\sqrt{1+a_n^2}$. The iteration of the process is prolonging the a_n side by the same length as the hypotenuse. This means, we append an isosceles triangle to our right-angle triangle. So, by an almost obvious angle computation, the angle opposite to the side of length 1 becomes half of what it was with each step. Since we start with $45^\circ = \frac{\pi}{4}$ at step1, the angle at step n is $\frac{\pi}{2^{n+1}}$. So

$$\lim_{n\to\infty} \frac{2^n}{a_n} = \lim_{n\to\infty} 2^n \tan\frac{\pi}{2^{n+1}} = \lim_{n\to\infty} \frac{2}{\pi} \cdot \frac{\tan\frac{\pi}{2^{n+1}}}{\frac{\pi}{2^{n+1}}} = \frac{2}{\pi}.$$

Since $\frac{\tan x}{x} \underset{x\to 0}{\to} 1$

Second solution. Denote $\alpha_n = \arctan a_n$, then $\tan\alpha_n = a_n$. Then

$$\tan(\alpha_{n+1}) = \tan\alpha_n + \sqrt{1+\tan^2\alpha_n} = \tan\alpha_n + \frac{1}{\cos\alpha_n} = \frac{\sin\alpha_n + 1}{\cos\alpha_n} = \frac{\cos\left(\frac{\pi}{2} - \alpha_n\right) + 1}{\sin\left(\frac{\pi}{2} - \alpha_n\right)} =$$

$$= \frac{2\cos^2\left(\frac{\pi}{4} - \frac{1}{2}\alpha_n\right)}{2\cos\left(\frac{\pi}{4} - \frac{1}{2}\alpha_n\right)\sin\left(\frac{\pi}{4} - \frac{1}{2}\alpha_n\right)} = \frac{\cos\left(\frac{\pi}{4} - \frac{1}{2}\alpha_n\right)}{\sin\left(\frac{\pi}{4} - \frac{1}{2}\alpha_n\right)} = \tan\left(\frac{\pi}{4} + \frac{1}{2}\alpha_n\right)$$

Therefore $\alpha_{n+1} = \frac{\pi}{4} + \frac{1}{2}\alpha_n$.

It is easier to consider take $\alpha_n = \frac{\pi}{2} - \beta_n$.

Then $\frac{\pi}{2} - \beta_{n+1} = \frac{\pi}{4} + \frac{\pi}{4} - \frac{1}{2}\beta_n$.

As $\beta_{n+1} = \frac{1}{2}\beta_n$.

Since $\alpha_1 = \arctan a_1 = \arctan 1 = \frac{\pi}{4}$, and $\beta_1 = \frac{\pi}{4}$ as well, hence $\beta_n = \dfrac{\pi}{2^{n+1}}$.

So $a_n = \tan \alpha_n = \cot \beta_n = \cot \dfrac{\pi}{2^{n+1}}$. We finish as in the first solution.

5. Polynomials $P(x)$ and $Q(x)$ of odd degree are such that for each integer x there is integer y such that $P(x) = Q(y)$. Prove that there exists a polynomial R, such that $P(x) = Q(R(x))$ for each x.

Remark. The condition of having odd degree is artificial. It makes problem technically much simpler and more suitable for a competition with limited time, but ideologically the same. We shall further comment regarding how to remove this restriction.

Solution. For $|x|$ large enough, both polynomial are monotone. We may assume WLOG that both P and Q have positive leading coefficient; indeed, if P has negative leading coefficient, we can replace P and Q by $-P$ and $-Q$, and if Q has negative leading coefficient, we can replace $Q(x)$ by $Q(-x)$.
It is enough to prove the equality for large x, since it is equality of polynomials. For large x, both P and Q are monotonically increasing, therefore the function $F = Q^{-1} \circ P$ in well-defined. It is an algebraic function, which receives integer values at integer points. By algebraic function we mean a function, Satisfying an equation of the form $a_n(x)F^n + a_{n-1}(x)F^{n-1} + \ldots + a_0(x) = 0$, where $a_i(x)$ are polynomials.
To separate the ideology of solution from technical details, we shall formulate several lemmas.

We shall say that $U(x) \prec V(x)$ if for sufficiently large x, $\left|\dfrac{U(x)}{V(x)}\right| < const$.

Lemma 1. $F(x) \prec Cx^{m/\ell}$, where $m = \deg P$, $\ell = \deg Q$.

Lemma 2. $F^{(s)}(x) \prec x^{\frac{m}{\ell} - s}$.

Define discrete derivative: $\Delta f(x) = f(x+1) - f(x)$. Discrete derivative can be applied several times, to obtain $\Delta^2 F = f(x+2) - 2f(x+1) + f(x)$, and so on; we will get formulas with alternating signs and binomial coefficients.

Lemma 3. For every n, there exists universal constant c_n, such that for every function f which has n continuous derivatives, and for each real x, it is possible to choose $y \in [x, x+n]$ such that $\Delta^{(n)} f(x) = c_n f^{(n)}(y)$.

Lemma 4. If for some n, $\Delta^{(n)} f(x) = 0$ for all integer x which is large enough, then f is a polynomial at sufficiently large integers (of degree less than n).

Using these lemmas, we can solve the problem easily. Indeed, choosing $s > \frac{m}{\ell} + 1$, we will have that $F^{(s)}(x) \underset{x\to\infty}{\to} 0$. Therefore, $\Delta^{(s)} F(x) \underset{x\to\infty}{\to} 0$, by lemma 3. However, F and hence $\Delta^{(s)} F$ are integer for sufficiently large integers, hence it is zero for sufficiently large integers, so by lemma 4, F is a polynomial for sufficiently large integers. Hence there is a polynomial R such that $P(x) = Q(R(x))$ at infinite number of points, hence it is true at all points (since nonzero polynomial cannot have infinite number of roots).

Now it remains to prove the lemmas, but first we shall hint about what problems can appear when we remove the condition of odd degree, and how to treat them. The point is, that when we define $F = Q^{-1} \circ P$ for sufficiently large x, sometimes we shall use values of Q at points far from zero, but not necessarily from the same side. It is better to define F_1 and F_2, one them will be integer for any sufficiently large integer x, but one tends to $+\infty$ and another to $-\infty$. This defines a way to paint sufficiently large integers in two colors. By Van der Waerden theorem, it is possible to choose an arbitrarily long monochromatic arithmetic sequences, and then the argument can be concluded in a similar way. We shall not explain the details of the general case here.

So, to formally complete the proof, we need to prove lemmas 1-4.

Proof of lemma 1. We shall write $f \sim g$, if there exist positive number c, C su ch that $c < \left|\frac{f}{g}\right| < C$ for sufficiently large x.

Then $x^m \sim P(x) = Q\big(F(x)\big) \sim \big(F(x)\big)^{\ell}$, hence $x^{m/\ell} \sim F(x)$.

Proof or lemma 2. The proof is by induction on s. The case $s = 0$ is lemma 1.

For the inductive step, differentiate s times the relation $Q(F(x)) = P(x)$.

We get a slightly terrifying expression of the form

$$\sum A_{k,t_1,\ldots,t_u} Q^{(k)}\big(F(x)\big) \cdot F^{(t_1)}(x) \cdot F^{(t_2)}(x) \cdot \ldots \cdot F^{(t_k)}(x) = P^{(s)}(x)$$

where A_* universal constants. In each summand, $t_1 + t_2 + \ldots + t_k = s$.

The only term that contains $F^{(s)}$ is $Q'\big(F(x)\big) \cdot F^{(s)}(x)$.

Notice that by induction $F^{(t)} \prec x^{\frac{m}{k}-t}$ for all $t < s$.

Hence $F^{(t_1)}(x) \cdot F^{(t_2)}(x) \cdot \ldots \cdot F^{(t_k)}(x) \prec x^{k\frac{m}{\ell}-t_1-t_2-\ldots-t_k} = x^{k\cdot\frac{m}{\ell}-s}$.

$Q^{(k)}$ is a polynomial of degree $\ell - k$, so $Q^{(k)}\big(F(x)\big) \sim F^{\ell-k} \sim x^{\frac{m}{\ell}(\ell-k)}$.

Hence each term

$$A_{k,t_1,\ldots,t_u} Q^{(k)}\big(F(x)\big) \cdot F^{(t_1)}(x) \cdot F^{(t_2)}(x) \cdot \ldots \cdot F^{(t_k)}(x) \prec x^{k\cdot\frac{m}{\ell}-s} \cdot x^{\frac{m}{\ell}(\ell-k)} = x^{m-s}.$$

Also, $P^{(s)} \sim x^{m-s}$.

If in the identity we move all terms except $Q'\big(F(x)\big) \cdot F^{(s)}(x)$ to the right hand side, we get

$$Q'\big(F(x)\big) \cdot F^{(s)}(x) \prec x^{m-s}.$$

But Q' is a polynomial of degree $s-1$, so $Q'(F) \sim F^{\ell-1} \sim x^{(\ell-1)\frac{m}{\ell}}$

$$x^{(\ell-1)\frac{m}{\ell}} \cdot F^{(s)}(x) \prec x^{m-s}$$

So $x^{(\ell-1)\frac{m}{\ell}} \cdot F^{(s)}(x) \prec x^{m-s-m+\frac{m}{\ell}} = x^{\frac{m}{\ell}-s}$. QED.

Proof of lemma 3. It is possible to choose such polynomial $p(x)$ of degree at most n, which has precisely the same values as f at points

$$x, x+1, ..., x+n.$$

The function $f-p$ has $n+1$ root in $[x, x+n]$, so by iteration of Rolle theorem, $(f-p)^{(n)}$ has at least one root y in $[x, x+n]$. Then $p^{(n)}(y) = f^{(n)}(y)$.

Then $p^{(n)}$ is a constant. It easy to see that Δ of a polynomial is a polynomial of degree 1 less. So $\Delta^{(n)} p$ is a constants, which depends linearly on the coefficient of x^n, as well as $p^{(n)}$. Hence $p^{(n)}(y) = f^{(n)}(y)$ is a universal constant times $\Delta^{(n)} p(x) = \Delta^{(n)} f(x)$.

Exercise to the reader. Compute this universal constant as a function of n ☺.

Proof of lemma 4. One shows inductively, that is Δp is polynomial of degree n for natural x, then p is a polynomial of degree $n+1$ for natural x. If you survived so far, you probably prefer to prove it yourself.

6. For given 2×2 matrices A, B there is only finite number n of 2×2 matrices X such that $X^2 + AX + B = 0$. Find the maximal possible value of n. (All matrices in this questions have complex entries.)

Solution. We shall denote $A = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$, $B = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix}$.

Let v be an eigenvector of X, $Xv = \lambda v$.

Then $X^2v + AXv + Bv = 0$, hence $\lambda^2 v + \lambda Av + Bv = 0$, so v is in the kernel of linear operator $P(\lambda) = \lambda^2 + \lambda A + B$. Therefore

$$\lambda^2 + \lambda A + B = \begin{pmatrix} \lambda^2 + a_1\lambda + b_1 & a_2\lambda + b_2 \\ a_3\lambda + b_3 & \lambda^2 + a_4\lambda + b_4 \end{pmatrix}$$

should be a degenerate matrix. Hence λ should be a root of the polynomial

$$p(\lambda) = \det(P(\lambda)) = \det(\lambda^2 + \lambda A + B) = \det\begin{pmatrix} \lambda^2 + a_1\lambda + b_1 & a_2\lambda + b_2 \\ a_3\lambda + b_3 & \lambda^2 + a_4\lambda + b_4 \end{pmatrix} =$$

$$= (\lambda^2 + a_1\lambda + b_1)(\lambda^2 + a_4\lambda + b_4) - (a_2\lambda + b_2)(a_3\lambda + b_3)$$

In the solution, we shall use a derivative of this polynomial, so we shall compute it now.

$p'(\lambda) =$

$$= (2\lambda + a_1)(\lambda^2 + a_4\lambda + b_4) - a_2(a_3\lambda + b_3) + (\lambda^2 + a_1\lambda + b_1)(2\lambda + a_4) - a_3(a_2\lambda + b_2)$$

We can open the brackets, but we won't.

If $P(\lambda_0)$ is a zero matrix, it means all entries of $P(\lambda)$ are divisible by $\lambda - \lambda_0$, so the $p(\lambda)$ is divisible by $(\lambda - \lambda_0)^2$. So, in this case λ_0 has to be a multiple root of $p(\lambda)$. Hence, if λ is a root of multiplicity 1 of p, then $P(\lambda)$ is not a zero matrix, hence the kernel of $P(\lambda)$ is one-dimensional.

Also, the eigenvector v of X has to be in the kernel of $\lambda^2 + \lambda A + B$. So if λ is a root of multiplicity 1 of p, an λ is an eigenvalue of X, then the direction of eigenvector v is defined uniquely.

(1) Assume that $p(\lambda)$ has no multiple roots. There are two cases: X can have distinct eigenvalues, or multiple eigenvalues. If X has distinct eigenvalues, they can be chosen in $\binom{4}{2}$ ways among the roots of $p(\lambda)$. Once we have chosen eigenvalues of X, the directions of eigenvectors are defined uniquely, and if eigenvectors are chosen, then X is defined uniquely in its eigenbasis. Hence each choice of two distinct eigenvalues of X among the roots of $p(\lambda)$ defines a unique matrix X, so there are 6 such matrices.

Now assume, that X has only one eigenvalue λ (of algebraic multiplicity 2). It also has to be a root of p. In this case $X=\lambda+N$, where N is a nilpotent matrix: $N^2=0$. Then

$$0=X^2+AX+B=(\lambda+N)^2+A(\lambda+N)+B=P(\lambda)+(2\lambda+A)N$$

$P(\lambda)$ is not a zero matrix, but it is degenerate, so it has one-dimensional kernel, which will be denoted K.

N also has a nontrivial kernel, and $(2\lambda+A)N$ has at least the same kernel, so N has to have the same kernel K. Being nilpotent, N specifies a mapping to its kernel, which is uniquely defined by specifying for a given vector outside K its image in K. So N is defined up to scaling, $N=sN_0$, where N_0 is a specific nilpotent matrix and s is a number. The condition that we have to satisfy is

$$0=P(\lambda)+(2\lambda+A)N=P(\lambda)+(2\lambda+A)sN_0$$

Is a linear condition in s. It either has an infinite number of solutions, or at most one solution. Let us multiply the last equation by the adjoint matrix of $2\lambda+A$.

$$0=\operatorname{adj}(2\lambda+A)\cdot P(\lambda)+\det(2\lambda+A)N$$

The second summand has trace zero. So the first also should have trace zero.
It is a necessary condition for existence of N. In coordinates

$$P(\lambda)=\begin{pmatrix}\lambda^2+a_1\lambda+b_1 & a_2\lambda+b_2\\ a_3\lambda+b_3 & \lambda^2+a_4\lambda+b_4\end{pmatrix},\quad 2\lambda+A=\begin{pmatrix}2\lambda+a_1 & a_2\\ a_3 & 2\lambda+a_4\end{pmatrix},$$

$$\operatorname{adj}(2\lambda+A)\cdot P(\lambda)=\begin{pmatrix}2\lambda+a_4 & -a_2\\ -a_3 & 2\lambda+a_1\end{pmatrix}\begin{pmatrix}\lambda^2+a_1\lambda+b_1 & a_2\lambda+b_2\\ a_3\lambda+b_3 & \lambda^2+a_4\lambda+b_4\end{pmatrix}=$$

$$=\begin{pmatrix}(2\lambda+a_4)(\lambda^2+a_1\lambda+b_1)-a_2(a_3\lambda+b_3) & *\\ * & (2\lambda+a_1)(\lambda^2+a_4\lambda+b_4)-a_3(a_2\lambda+b_2)\end{pmatrix}$$

We have computed only diagonal elements, since we aim to compute trace. So

$$0=\operatorname{tr}\left(\operatorname{adj}(2\lambda+A)\cdot P(\lambda)\right)=(2\lambda+a_4)(\lambda^2+a_1\lambda+b_1)-a_2(a_3\lambda+b_3)+$$
$$+(2\lambda+a_1)(\lambda^2+a_4\lambda+b_4)-a_3(a_2\lambda+b_2)=p'(\lambda)$$

But in our case, when p has no roots with multiplicity greater than one, there are no common roots for p and p', so there are no such solutions.

So, in this case, we have only 6 possible values for X.

(2) Assume that p has a root of multiplicity greater than 1. In this case, p has at most 3 distinct roots. We shall distinguish two cases, depending on whether there exists λ such that $P(\lambda)$ is a zero matrix.

(2.0) If there exists λ_0, such that $P(\lambda_0)$ is a zero matrix. Assume there is also $\lambda_1 \neq \lambda_0$ such that $p(\lambda_1)=0$. There is a nonzero vector v_1, such that $P(\lambda_1)\cdot v_1 = 0$. For any vector v_0 we have $P(\lambda_0)\cdot v_0 = 0$. If we choose arbitrary v_0, which is not multiple of v_1, then there is a unique matrix X such that $Xv_0 = \lambda_0 v_0$ and $Xv_1 = \lambda_1 v_1$, and since there is infinite number of ways to choose the direction of v_0, the are infinite number of ways to construct such X. In all cases, $P(X)$ is zero on v_0 and v_1, and hence it is a zero matrix. So there are infinite number of solutions, which is forbidden.

Now assume that all four roots of p are equal to the same value λ_0. In this case all eigenvalues of X are λ_0. Then $X = \lambda_0 + N$, where N is nilpotent. So

$$P(X) = (\lambda_0 + N)^2 + A(\lambda_0 + N) + B = P(\lambda_0) + (2\lambda_0 + A)N = (2\lambda_0 + A)N = 0.$$

If we have at least one option for N which is not a zero matrix, then for each number μ, also μN works. Hence there is either infinite number of solutions (which is forbidden) or there is just one solution in this case.

(2.1) Now we assume that p has a multiple root, but $P(\lambda)$ is never a zero matrix. Than we have at most 3 different roots, so we can choose two distinct roots $\lambda_1 \neq \lambda_2$ in at most 3 ways. Matrices $P(\lambda_1)$ and $P(\lambda_2)$ are degenerate but non-zero, so a non-zero vectors $v_1 \in \ker P(\lambda_1)$, $v_2 \in \ker P(\lambda_2)$ are defined uniquely up to scaling, Therefore a matrix X such that $Xv_i = \lambda_i v_i$ for $i = 1,2$ is unique. So there are only 3 solutions with distinct eigenvalues.

Assume that X has just one eigenvalue (of algebraic multiplicity 2). This eigenvalue λ can be chosen in 3 possible ways. Assume we have chosen λ, then $X = \lambda + N$, where N is a nilpotent matrix, and therefore

$$0 = P(X) = P(\lambda + N) = P(\lambda) + (2\lambda + A)N.$$

Now $P(\lambda)$ is a degenerate but nonzero matrix, so as in part (1), N has the same kernel as $P(\lambda)$, and hence $N = sN_0$, so where N_0 is a specific nilpotent matrix and s is a number. The equation is linear in s, so it has either infinite number of solutions (which is forbidden), or at most one solution.

So, in this case (with multiplicities) we get at most 3 solutions with distinct eigenvalues, and at most 3 solutions with double eigenvalues, so in total at most 6 solutions.

Remark. In this case, the estimate can be improved, but there's no need to.

An example for having precisely 6 solutions: $P(X) = X^2 + \begin{pmatrix} 0 & 10 \\ 1 & 0 \end{pmatrix} X + \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$.

$$p(\lambda) = \det\begin{pmatrix} \lambda^2+1 & 10\lambda \\ \lambda & \lambda^2+4 \end{pmatrix} = (\lambda^2+1)(\lambda^2+4) - 10\lambda^2 = \lambda^4 + 5\lambda^2 + 4 - 10\lambda^2 =$$

$$= \lambda^4 - 5\lambda^2 + 4 = (\lambda^2-1)(\lambda^2-4) = (\lambda-1)(\lambda+1)(\lambda-2)(\lambda+2)$$

So for arbitrary choice of $\lambda_1 \neq \lambda_2$ from the set $\{-2,-1,1,2\}$, we can find non-zero vectors $v_1 \in \ker P(\lambda_1)$ and $v_2 \in \ker P(\lambda_2)$. It is easy to see that v_1, v_2 are linearly independent. Indeed, if there is a vector v_0 in $\ker P(\lambda_1) \cap \ker P(\lambda_2)$ then

$$0 = (P(\lambda_1) - P(\lambda_2))v_0 = \left(\lambda_1^2 - \lambda_2^2 + (\lambda_1 - \lambda_2)\begin{pmatrix} 0 & 10 \\ 1 & 0 \end{pmatrix}\right)v =$$

$$= (\lambda_1 - \lambda_2)\left(\lambda_1 + \lambda_2 + \begin{pmatrix} 0 & 10 \\ 1 & 0 \end{pmatrix}\right)v = (\lambda_1 - \lambda_2)\begin{pmatrix} \lambda_1+\lambda_2 & 10 \\ 1 & \lambda_1+\lambda_2 \end{pmatrix}v$$

So either $\lambda_1 = \lambda_2$, or $v = 0$, or $\det\begin{pmatrix} \lambda_1+\lambda_2 & 10 \\ 1 & \lambda_1+\lambda_2 \end{pmatrix} = (\lambda_1+\lambda_2)^2 - 10 = 0$.

The last possibility doesn't exist, since summing numbers from $\{-2,-1,1,2\}$ won't produce $\pm\sqrt{10}$. Therefore, we can construct a unique matrix satisfying $Xv_i = \lambda_i v_i$ for $i = 1, 2$, so in this case we have precisely 6 solutions (and we don't have more, because of the discussion of case (1).