I think that your remark is interesting. Indeed we can reformulate the initial problem:

$$\dot{\mathbf{x}} = \frac{p(\mathbf{x})}{q(\mathbf{x})}$$

as

$$\frac{dQ(\mathbf{x})}{dt} = p(\mathbf{x})$$

But then we end up with learnable coefficients on both sides of the equality and I am not sure if this is something we want. Also we would still need to compute the time-derivative of Q which would require to go back to the precedent formulation. Maybe I am missing out something but I really don't see how to reformulate this into a least squares problem.

What seems more natural to me is to treat this as a non-linear least squares problem and instead of having:

$$\dot{X} = \theta(X) \times \xi$$

that we had so far with the polynomial candidate functions, this was a linear least squares (with the same notations as in our slides).

We could say that now we have:

$$\dot{X} = \tilde{g}(X, \xi) \text{ or equivalently } \begin{bmatrix} \dot{x}(t_1) \\ \dot{x}(t_2) \\ \vdots \\ \dot{x}(t_m) \end{bmatrix} = \begin{bmatrix} g(x(t_1), \xi) \\ g(x(t_2), \xi) \\ \vdots \\ g(x(t_m), \xi) \end{bmatrix}$$

where

$$g(x, \xi) = \frac{p_\xi(x)}{q_\xi(x)}$$

Where ξ are the learnable parameters of our problem which means that they are the coefficients of the polynomials p_ξ and q_ξ , $\tilde{g}$ is the vectorized version of g. Here we would minimize the $||\dot{X} - \tilde{g}(X, \xi)||_2^2$ with respect to ξ