1. (Why on page 2 of the November 24 slides there is $x_k = (A - tI)^{-1} q_k$ earlier and $q_k = (A - tI)^{-1} x_k$ later) You are correct - I switched x_k and q_k in the second formula. The correct version is the one in the pseudocode.

2. (Is it correct that when using the non-normalized power method $x_k \to \infty$ when $|\lambda_1| > 1$) You are correct: the typical behavior for the power method when $|\lambda_1| > 1$ is that the entries of x_k diverge. Two things happen: x_k gets longer as k increases, and it also gets aligned with the direction of the leading eigenvector v_1. This is discussed in the November 23 lecture, starting from page 4: first I draw an example of this typical behavior in $\mathbb{R}^2$, then we prove a more precise assertion:

   Lemma 1. *If A is diagonalizable (i.e., $A = VDV^{-1}$ with D diagonal), whenever x_0 is chosen such that the first entry of $V^{-1}x_0$ is nonzero, then $\frac{1}{\gamma_k} x_k$ tends to a multiple of the eigenvector v_1 (where $\gamma_k = \lambda_1^k$).*

   This is proved first in the case in which A is diagonal and then in page 6 we extend the proof to a diagonalizable A.

   In particular, the fact that $\frac{1}{\gamma_k} x_k \to v_1$ implies that for each index $h = 1, 2, ..., n$ such that the entry $(v_1)_h$ of v_1 is nonzero, the corresponding entry $(x_k)_h$ diverges:

   $$\lim (x_k)_h = \lim \left(\frac{1}{\gamma_k} x_k \right)_h \cdot \lim \gamma_k = (v_1)_h \cdot \infty = \infty.$$

   (Note that the last steps works only if $(v_1)_h \neq 0$.)

3. (In subspace iteration, why can we ignore factors that post-multiply Z_k (or X_k) — I hope I understand this question correctly.) The main reason is the following result: let $\operatorname{Im} M$ denote the column space of a matrix M (also called image, or range; i.e., the space spanned by all its columns).

   Lemma 2. *Let $X, Z \in \mathbb{C}^{n \times p}$, with $n \geq p$, be such that $X = ZR$, with $R \in \mathbb{C}^{p \times p}$ nonsingular. Then, $\operatorname{Im} X = \operatorname{Im} Z$.*

   Proof. Let $X_h = Xe_h$ be the hth column of X (and similarly for R_h and Z_h): then, we have for each h

   $$X_h = (ZR)_h = Z_1 R_{1h} + Z_2 R_{2h} + \cdots + Z_k R_{ph}.$$

   This shows that $X_h \in \operatorname{Im} Z$, because it is a linear combination of the columns of Z. Since this holds for each h, we have $\operatorname{Im} X \subseteq \operatorname{Im} Z$.

   We have $Z = XR^{-1}$, so we can repeat the same argument swapping X and Z, with R^{-1} instead of R, and prove that $\operatorname{Im} Z \subseteq \operatorname{Im} X$. $\square$

   So if we are only interested in the column space of a matrix X, we are free to replace it with itself multiplied (to the right) by a square invertible matrix.

We use this lemma in several places: when we implement subspace iteration, we replace at each step X_k with $\hat{Z}_k = X_k(\hat{R}_k)^{-1}$. In particular, if we start from the same $X_0 = \hat{Z}_0$, the sequences X_k (defined by $X_k = AX_{k-1}$ or $X_k = A^k X_0$, hence obtained without 'replacements') and $\hat{Z}_k$ (defined by $[\hat{Z}_k, R_k] = \texttt{thin\_qr}(A\hat{Z}_{k-1})$, so with 'replacements' at each step) have the same image: indeed,

$$X_1 = AX_0 = \hat{Z}_1\hat{R}_1, \tag{1a}$$
$$X_2 = AX_1 = A\hat{Z}_1\hat{R}_1 = Z_2\hat{R}_2\hat{R}_1, \tag{1b}$$
$$X_3 = AX_2 = A\hat{Z}_2\hat{R}_2\hat{R}_1 = \hat{Z}_3\hat{R}_3\hat{R}_2\hat{R}_1, \tag{1c}$$
$$\dots \quad \dots \tag{1d}$$

so at each step X_k is equal to $\hat{Z}_k$ multiplied by the matrix $\hat{R}_k\hat{R}_{k-1}\cdots\hat{R}_2\hat{R}_1$ (this argument can be easily turned into an induction proof).

Similarly, when we analyze the method, we make a QR factorization $X_0^* = QS^*$, or, transposing, $X_0 = SQ^*$, and then we focus on computing the image of A^kS rather than A^kX_0, because they only differ by multiplication by the invertible Q^*.

(In the November 24 lecture I did not give an explicit name to S, but I just replaced X_0 with $X_0Q = S$ at some point; I am sorry if this made things more confusing.)

(In the beginning of the November 29 lecture, I tried to re-explain better this part on why we can multiply by an invertible matrix on the right.)

There is one last thing that you may be asking yourself about at this point, and it is 'for the lemma to work, you need the $\hat{R}_k$ to be invertible at each step; why is this the case?'. The answer is that $\hat{R}_k$ is not always invertible (for instance, if $A = 0$, no matter what X_0 we choose, $AX_0 = 0$, and so we must have $\hat{R}_1 = 0$ in the thin QR) but it is invertible under the hypotheses that we made in our convergence analysis of the method. For completeness, here is a proof.

Lemma 3. *Let A be diagonal, and suppose that the square submatrix formed by the first p rows of $X_0 = \hat{Z}_0$ is invertible. Then, at each step k of the subspace iteration,*

- *the submatrix formed by the first p rows of X_k is invertible;*
- *the submatrix formed by the first p rows of $\hat{Z}_k$ is invertible;*
- *$\hat{R}_k$ is invertible.*

Proof. For ease of notation, for a matrix $M \in \mathbb{C}^{n\times p}$ we call $F(M)$ the $p \times p$ matrix formed by its first p rows.

As in the Nov 24 lecture, we make the thin QR factorization $X_0^* = QS^*$, or

$$X_0 = SQ^* = \begin{bmatrix} x_{11} & 0 & 0 & \dots & 0 \\ x_{21} & x_{22} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ x_{p-1,1} & x_{p-1,2} & \dots & x_{p-1,p-1} & 0 \\ x_{p1} & x_{p2} & \dots & \dots & x_{pp} \\ x_{p+1,1} & x_{p+2,2} & \dots & \dots & x_{p+1,p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & \dots & x_{np} \end{bmatrix} Q^*$$

If we take only the leading $p \times p$ block of both sides of this equality, we have the equation $F(X_0) = F(S)Q^*$. Here $F(X_0)$ and Q^* are invertible, so $F(S)$ must be invertible too: to see this, you can take determinants, for instance: $0 \neq \det F(X_0) = \det(F(S)Q^*) = \det F(S) \det Q^*$, so none of the determinants in the right-hand side can be zero.

In particular, we have $x_{11}, x_{22}, \dots, x_{pp} \neq 0$ (a triangular matrix is nonsingular if and only if its diagonal entries are nonzero).

It follows from the computation in our Nov 24 lecture (page 6 in the handwritten notes) that

$$F(A^k S) = \begin{bmatrix} \lambda_1^k x_{11} & 0 & 0 & \dots & 0 \\ \lambda_2^k x_{21} & \lambda_2^k x_{22} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda_{p-1}^k x_{p-1,1} & \lambda_{p-1}^k x_{p-1,2} & \dots & \lambda_{p-1}^k x_{p-1,p-1} & 0 \\ \lambda_p^k x_{p1} & \lambda_p^k x_{p2} & \dots & \dots & \lambda_p^k x_{pp} \end{bmatrix},$$

and this matrix is invertible, too (because it is a triangular matrix with nonzeros on the diagonal). So also $F(X_k) = F(A^k X_0) = F(A^k S Q^*) = F(A^k S)Q^*$ is invertible. In view of the Equations (1) above in this note, $F(X_k) = F(\hat{Z}_k \hat{R}_k \hat{R}_{k-1} \dots \hat{R}_1) = F(\hat{Z}_k)\hat{R}_k \hat{R}_{k-1} \dots \hat{R}_1$, and this implies that $F(\hat{Z}_k), \hat{R}_k, \hat{R}_{k-1}, \dots, \hat{R}_1$ must all be nonsingular. (Again, you can take determinants to see it.) $\square$

Again, we can generalize from the case of A diagonal to the one of A diagonalizable:

Lemma 4. *Let A be diagonalizable (i.e., $A = VDV^{-1}$ with D diagonal), and suppose that the square submatrix formed by the first p rows of $V^{-1}X_0 = V^{-1}\hat{Z}_0$ is invertible. Then, at each step k of the subspace iteration,*

- *the submatrix formed by the first p rows of $V^{-1}X_k$ is invertible;*
- *the submatrix formed by the first p rows of $V^{-1}\hat{Z}_k$ is invertible;*
- *$\hat{R}_k$ is invertible.*

(Note that the condition on the $p \times p$ block of $V^{-1}X_0$ is a generalization of the condition on the first entry of $V^{-1}x_0$ in the power method.)

Demmel's book deals with orthogonal iteration in Section 4.4.3, but he only states without proof that $\operatorname{Im} \hat{Z}_i = \operatorname{Im}(A^i \hat{Z}_0)$, and then makes some arguments about convergence of subspaces (which I find less convincing than the version I explained to you – that is why I chose to go through a different route.)

4. (How can one prove that X_k and $\hat{Z}_k$ span the same subspace) I have already answered to this in the previous point: the computation around Equations 1 shows that $X_k = \hat{Z}_k \hat{R}_k \hat{R}_{k-1} \cdots \hat{R}_1$; Lemma 4 shows that under our suitable hypotheses $\hat{R}_k$ is invertible for each k, and Lemma 2 shows that in this case $\hat{Z}_k$ and X_k span the same subspace.