Consider the problem:

$$\min_{\boldsymbol{x}} \frac{\mu}{2}\|\boldsymbol{W}^T\boldsymbol{Q}\boldsymbol{W}\boldsymbol{x} - \boldsymbol{W}^T\boldsymbol{Q}\boldsymbol{W}\boldsymbol{s}\|_2^2 + \frac{\lambda}{2}\sum_i \|\boldsymbol{G}_i\boldsymbol{x}\|_2^2 \tag{1}$$

where $\boldsymbol{Q}$ is a nonlinear quantization operator such that $\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}$ implies $\boldsymbol{y}[i] = \boldsymbol{q}[i] * \text{round}(\frac{\boldsymbol{x}[i]}{\boldsymbol{q}[i]})$.

I can rewrite this problem, splitting $\boldsymbol{x}$ into two variables:

$$\begin{aligned} &\arg\min_{\boldsymbol{x},\boldsymbol{z}} \frac{\mu}{2}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\boldsymbol{G}_i\boldsymbol{x}\|_2^2 \\ &\text{subject to } \boldsymbol{Q}\boldsymbol{W}\boldsymbol{z} - \boldsymbol{Q}\boldsymbol{W}\boldsymbol{s} = 0 \end{aligned} \tag{2}$$

It would certainly be possible to swap the constraint and the first objective term, but I would like to minimize the number of update equations for which the nonlinear operator $\boldsymbol{Q}$ complicates.

The alternating directionn method of multipliers algorithm (ADMM) is designed for affine equality constrains, but $\boldsymbol{W}^T\boldsymbol{Q}\boldsymbol{W}$ is close enough to linear that I expect it will succeed here despite the nonlinearity of quantization.

1 $\boldsymbol{x}$ Update

2 $\boldsymbol{z}$ Update

$$\boldsymbol{z}^{(k+1)} = \arg\min_{\boldsymbol{z}} \frac{\mu}{2}\|\boldsymbol{z}-\boldsymbol{x}^{(k+1)}\|_2^2+\frac{\rho}{2}\|\boldsymbol{W}^T\boldsymbol{Q}\boldsymbol{W}\boldsymbol{z}+(1-\alpha)\boldsymbol{W}^T\boldsymbol{Q}\boldsymbol{W}\boldsymbol{z}^{(k)}-(2-\alpha)\boldsymbol{W}^T\boldsymbol{Q}\boldsymbol{W}\boldsymbol{s}+\frac{\gamma^{(k)}}{\rho}\|_2^2 \tag{3}$$

I should note here that due to the quantization, this minimizer is not guarenteed (or even likely) to exist. However, the infimum does exist and a solution that evaluates arbitrarily close to that infimum can be found.

To solve this, I will start by ignoring the quantization and finding the assymptotic solution as $\boldsymbol{q}$ approaches zero.

$$\boldsymbol{z}^{(k+1)} \approx \arg\min_{\boldsymbol{z}} \frac{\mu}{2}\|\boldsymbol{z}-\boldsymbol{x}^{(k+1)}\|_2^2+\frac{\rho}{2}\|\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{z}+(1-\alpha)\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{z}^{(k)}-(2-\alpha)\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{s}+\frac{\gamma^{(k)}}{\rho}\|_2^2 \tag{4}$$

$$(\mu\mathbf{I} + \rho\boldsymbol{W}^T\boldsymbol{W})\boldsymbol{z}^{(k+1)} \approx \mu\boldsymbol{x}^{(k+1)} + \rho\boldsymbol{W}^T((2-\alpha)\boldsymbol{W}\boldsymbol{s} - (1-\alpha)\boldsymbol{W}\boldsymbol{z}^{(k)} - \frac{\gamma}{\rho}) \tag{5}$$

$$\mu\mathbf{I} + \rho\boldsymbol{W}^T\boldsymbol{W} = \mu(\mathbf{I} - \boldsymbol{W}^T\boldsymbol{W}) + (\mu+\rho)\boldsymbol{W}^T\boldsymbol{W} \tag{6}$$

$$(\mu\mathbf{I} + \rho\boldsymbol{W}^T\boldsymbol{W})^{-1} = \frac{1}{\mu}(\mathbf{I} - \boldsymbol{W}^T\boldsymbol{W}) + \frac{1}{\mu+\rho}\boldsymbol{W}^T\boldsymbol{W} \tag{7}$$

$$(\mu \mathbf{I} + \rho \boldsymbol{W}^T \boldsymbol{W})^{-1} = \frac{1}{\mu}\mathbf{I} - \frac{\rho}{\mu^2 + \rho\mu}\boldsymbol{W}^T \boldsymbol{W} \tag{8}$$

$$\boldsymbol{z}^{(k+1)} \approx \boldsymbol{x}^{(k+1)} - \frac{\rho}{\mu + \rho}\boldsymbol{W}^T \boldsymbol{W} \boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu + \rho}\boldsymbol{W}^T((2-\alpha)\boldsymbol{W}\boldsymbol{s} - (1-\alpha)\boldsymbol{W}\boldsymbol{z}^{(k)} - \frac{\gamma}{\rho}) \tag{9}$$

$$\boldsymbol{z}^{(k+1)} \approx \boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu + \rho}\boldsymbol{W}^T((2-\alpha)\boldsymbol{W}\boldsymbol{s} - \boldsymbol{W}\boldsymbol{x}^{(k+1)} - (1-\alpha)\boldsymbol{W}\boldsymbol{z}^{(k)} - \frac{\gamma}{\rho}) \tag{10}$$

Now, I will reinsert $\boldsymbol{Q}$ in the places it was removed. The result should still be fairly close to the actual solution.

$$\boldsymbol{z}^{(k+1)} \approx \boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu + \rho}\boldsymbol{W}^T((2-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{s} - \boldsymbol{W}\boldsymbol{x}^{(k+1)} - (1-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{z}^{(k)} - \frac{\gamma}{\rho}) \tag{11}$$

In the interest of brevity, I will define:

$$\boldsymbol{r} = (2-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{s} - \boldsymbol{W}\boldsymbol{x}^{(k+1)} - (1-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{z}^{(k)} - \frac{\gamma}{\rho} \tag{12}$$

$$\boldsymbol{z}^{(k+1)}_{\text{approx}} = \boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu + \rho}\boldsymbol{W}^T \boldsymbol{r} \tag{13}$$

Plugging the $\boldsymbol{z} = \boldsymbol{z}^{(k+1)}_{\text{approx}} + \Delta \boldsymbol{z}$ into the function to be minimized, I have the expression:

$$\frac{\mu}{2}\|\frac{\rho}{\mu + \rho}\boldsymbol{W}^T\boldsymbol{r} + \Delta\boldsymbol{z}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{Q}\boldsymbol{W}(\boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu+\rho}\boldsymbol{W}^T\boldsymbol{r} + \Delta\boldsymbol{z}) - (2-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{s} + (1-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{z}^{(k)} + \frac{\gamma^{(k)}}{\rho}\|_2^2 \tag{14}$$

$$\frac{\mu}{2}\|\frac{\rho}{\mu + \rho}\boldsymbol{W}^T\boldsymbol{r} + \Delta\boldsymbol{z}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{Q}\boldsymbol{W}(\boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu+\rho}\boldsymbol{W}^T\boldsymbol{r} + \Delta\boldsymbol{z}) - \boldsymbol{r} - \boldsymbol{W}\boldsymbol{x}^{(k+1)}\|_2^2 \tag{15}$$

$$\boldsymbol{W}\boldsymbol{W}^T = \mathbf{I} \tag{16}$$

$$\frac{\mu}{2}\|\frac{\rho}{\mu + \rho}\boldsymbol{W}^T\boldsymbol{r} + \Delta\boldsymbol{z}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{Q}(\boldsymbol{W}\boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu+\rho}\boldsymbol{r} + \boldsymbol{W}\Delta\boldsymbol{z}) - \boldsymbol{r} - \boldsymbol{W}\boldsymbol{x}^{(k+1)}\|_2^2 \tag{17}$$

Any component of $\Delta\boldsymbol{z}$ that is orthogonal to $\boldsymbol{W}$ will increase the first term of the objective expression and leave the second term unchanged. Therefore, the optimal choice of $\Delta\boldsymbol{z}$ has no component orthogonal to $\boldsymbol{W}$.

So, I can minimize this expression instead:

$$\frac{\mu}{2}\|\frac{\rho}{\mu + \rho}\boldsymbol{r} + \boldsymbol{W}\Delta\boldsymbol{z}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{Q}(\boldsymbol{W}\boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu+\rho}\boldsymbol{r} + \boldsymbol{W}\Delta\boldsymbol{z}) - \boldsymbol{r} - \boldsymbol{W}\boldsymbol{x}^{(k+1)}\|_2^2 \tag{18}$$

Finally, again for convenience, define

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}^{(k+1)} + \frac{\rho}{\mu+\rho}\boldsymbol{r} \tag{19}$$

So, I have

$$\frac{\mu}{2}\|\frac{\rho}{\mu+\rho}\boldsymbol{r} + \boldsymbol{W}\Delta\boldsymbol{z}\|_2^2 + \frac{\rho}{2}\| - \frac{\mu}{\mu+\rho}\boldsymbol{r} + \boldsymbol{Q}(\boldsymbol{y} + \boldsymbol{W}\Delta\boldsymbol{z}) - \boldsymbol{y}\|_2^2 \tag{20}$$

It is convenient that when trying to minimize the expression, each element of $\boldsymbol{W}\Delta\boldsymbol{z}$ can be treated independently.

In comparison to $\boldsymbol{W}\Delta\boldsymbol{z} = 0$, it is possible to decrease the first objective term while not changing the second objective term by adjusting $\boldsymbol{W}\Delta\boldsymbol{z}$ in the direction $-\boldsymbol{r}$.

$$\boldsymbol{W}\Delta\boldsymbol{z}[i] = -\operatorname{sign}(\boldsymbol{r}[i])\operatorname{minimum}(|\frac{\rho}{\mu+\rho}\boldsymbol{r}[i]|, |\boldsymbol{q}[i]*\operatorname{round}(\frac{\boldsymbol{y}[i]}{\boldsymbol{q}[i]}) - 0.5^{-}\boldsymbol{q}[i]\operatorname{sign}(\boldsymbol{r}[i]) - \boldsymbol{y}[i]| \tag{21}$$

where 0.5^{-} is a number less than 0.5, but arbitarily close to it.

In comparison to $\boldsymbol{W}\Delta\boldsymbol{z} = 0$, it may be possible to decrease the second term by using $\boldsymbol{W}\Delta\boldsymbol{z}$ to switch the rounding direction.

$$\boldsymbol{W}\Delta\boldsymbol{z}[i] = \boldsymbol{q}[i]*\operatorname{round}(\frac{\boldsymbol{y}[i]}{\boldsymbol{q}[i]}) - 0.5^{+}\boldsymbol{q}[i]\operatorname{sign}(\boldsymbol{y}[i] - \boldsymbol{q}[i]*\operatorname{round}(\frac{\boldsymbol{y}[i]}{\boldsymbol{q}[i]}) - \frac{\rho}{\mu+\rho}\boldsymbol{r}[i]) - \boldsymbol{y}[i] \tag{22}$$

Finally, in comparison to $\boldsymbol{W}\Delta\boldsymbol{z} = 0$, it may be possible to decrease the first objective term while switching the rounding direction.

$$\boldsymbol{W}\Delta\boldsymbol{z}[i] = -\operatorname{sign}(\boldsymbol{r}[i])\operatorname{minimum}(|\frac{\rho}{\mu+\rho}\boldsymbol{r}[i]|, |\boldsymbol{q}[i]*\operatorname{round}(\frac{\boldsymbol{y}[i]}{\boldsymbol{q}[i]}) - 1.5^{-}\boldsymbol{q}[i]\operatorname{sign}(\boldsymbol{r}[i]) - \boldsymbol{y}[i]| \tag{23}$$

By comparing these three possible minimizers[1], I can identify the minimizer, and adjust the $\boldsymbol{z}_{\text{approx}}^{(k+1)}$ accordingly.

$$\boldsymbol{z}^{(k+1)} = \boldsymbol{z}_{\text{approx}}^{(k+1)} + \Delta\boldsymbol{z} \tag{24}$$

3 γ update

$$\frac{\gamma^{(k+1)}}{\rho} = \frac{\gamma^{(k)}}{\rho} + \boldsymbol{Q}\boldsymbol{W}\boldsymbol{z}^{(k+1)} + (1-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{z}^{(k)} - (2-\alpha)\boldsymbol{Q}\boldsymbol{W}\boldsymbol{s} \tag{25}$$

[1] Again, I should clarify, not an actual minimizer, but something that evaluates arbitarily close to the infimum.