Supplementary Materials

I. Derivation of Equation (14)-(16).

Derivation from equation (14) to (15). Recall that we assumed an uniform prior distribution of $\mu$ as follows

$$p(\mu) = \begin{cases} 1/c & \text{if } \mu \in U, \\ 0 & \text{otherwise}, \end{cases}$$

(1)

where $U$ is a sufficiently large bounded set that includes $0$ and $c$ is the measure of $U$. For simplicity, we consider $U = [-U, U]^J$, where $J$ is the dimension of $\mu$. On the other hand, $q(\mu | \alpha) = \prod_j q(\mu_j | \alpha_j)$, where $q(\mu_j | \alpha_j)$ is the Bernoulli distribution with probability $q_j$.

By definition, we have

$$KL(\mu | \alpha) = \int q(\mu | \alpha) \log q(\mu | \alpha) d\mu - \int q(\mu | \alpha) \log p(\mu) d\mu.$$  

(2)

Note that $p(\mu)$ is a continuous distribution while $q(\mu | \alpha)$ is discrete. For the discrete Bernoulli distribution, $\int q(\mu | \alpha) \log q(\mu | \alpha) d\mu$ is calculated by

$$\int q(\mu | \alpha) \log q(\mu | \alpha) d\mu = \sum_j (q_j \log q_j + (1 - q_j) \log(1 - q_j)).$$

(3)

For the calculation of $\int q(\mu | \alpha) \log p(\mu) d\mu$, we use a continuous representation of the Bernoulli distribution as follows

$$q(\mu_j | \alpha_j) = (1 - q_j)D(\mu_j) + q_j D_{\alpha_j}(\mu_j),$$

(4)

where $D_\alpha(x)$ is the Dirac delta function that satisfies

$$\int D_\alpha(x) dx = 1.$$  

(5)

Next we consider two cases: $\alpha \in U$ and $\alpha \notin U$.

- $\alpha \in U$: It means that $\alpha_j \in [-U, U], \forall j$. Then

$$\int q(\mu | \alpha) \log p(\mu) d\mu = -\log c \prod_{q_j, \mu_j \in \{0, \alpha_j\}} \left( q_j I_{a_j}(\mu_j) + (1 - q_j) I_0(\mu_j) \right),$$

(6)

where $I_a(x)$ is the indicator function:

$$I_a(x) = \begin{cases} 1 & \text{if } x = a, \\ 0 & \text{if } x \neq a. \end{cases}$$

(7)

That is, $\int q(\mu | \alpha) \log p(\mu) d\mu$ is a constant regardless of the value of $\alpha$ as long as $\alpha \in U$.

- $\alpha \notin U$: it is direct to get that

$$\int q(\mu | \alpha) \log p(\mu) d\mu \leq \log p(\alpha) = -\infty.$$  

(8)

In summary, $KL(q(\mu | \alpha) || p(\mu)))$ is a constant if $\alpha \in U$ and $+\infty$ otherwise. Thus, we obtain

$$\min_\alpha KL(q(\mu | \alpha) || p(\mu)) - E_{\mu \sim q(\mu | \alpha)} \log p(y_0 | \mu) \leftrightarrow \max_\alpha E_{\mu \sim q(\mu | \alpha)} \log p(y_0 | \mu) - \delta_U(\alpha)$$

$$\leftrightarrow \max_\alpha E_{\mu \sim q(\mu | \alpha)} \log p(y_0 | \mu) \text{.}$$

(9)

Derivation from equation (15) to (16). Since we assume $x_0 = G(\epsilon_0)$ and $y_0 = Ax_0 + n$, where $n \sim N(0, \sigma^2 I)$, it can be obtained that $y_0 | \mu \sim \mathcal{N}(A G(\mu(0)), \sigma^2 I)$, i.e. $\log p(y_0 | \mu) \propto -||y_0 - AG(\mu(0))||_2^2$. Recall that $\mu = \alpha \odot d$, then we have

$$\max_\alpha E_{\mu \sim q(\mu | \alpha)} \log p(y_0 | \mu) \leftrightarrow \min_\alpha E_{\mu \sim q(\mu | \alpha)} ||y_0 - AG(\mu(0))||_2^2$$

$$\leftrightarrow \min_\alpha E_d ||y_0 - AG(\mu(0))||_2^2.$$  

(10)

The constraint $\alpha \in U$ is omitted in (16) as the feasible set $U$ is sufficiently large.
II. Test Sets

Fig. 1: Images of Natural-6 test set.

Fig. 2: Images of Unnatural-6 test set.

Fig. 3: Images of Set-20 test set.
III. ADDITIONAL VISUAL RESULTS

Fig. 4: Reconstructions on image “Barbara” in the presence of AWGN (SNR=15dB) with $R = 1$ bipolar mask.

Fig. 5: Reconstructions on image “House” in the presence of AWGN (SNR=20dB) with $R = 1$ bipolar mask.
Fig. 6: Pixel-wise statistics over 100 predictions on image “Ball”, “House”, “Butterfly”, and “Peppers”. Four bipolar masks are used in these experiments.
Fig. 7: Reconstructions on image “Cameraman” in the presence of Poisson noise ($\alpha = 27$) with $R = 2$ bipolar masks.

Fig. 8: Reconstructions on image “Flinstones” in the presence of Poisson noise ($\alpha = 9$) with $R = 4$ bipolar masks.
<table>
<thead>
<tr>
<th>Ground-Truth / PSNR</th>
<th>WF / 22.23dB</th>
<th>DOLPHIn / 23.80dB</th>
<th>BM3D-prGAMP / 22.39dB</th>
<th>ConPR / 22.48dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>prDeep / 27.13dB</td>
<td>DSR / 22.49dB</td>
<td>DIP / 23.42dB</td>
<td>Net-PGD / 19.32dB</td>
<td>DeepMMSE / 27.29dB</td>
</tr>
</tbody>
</table>

Fig. 9: Reconstructions on image “Pillars of Creation” from $4 \times$ oversampled Fourier magnitude measurements with Poisson noise ($\gamma = 2$).

<table>
<thead>
<tr>
<th>Ground-Truth / PSNR</th>
<th>WF / 17.30dB</th>
<th>DOLPHIn / 21.50dB</th>
<th>BM3D-prGAMP / 23.34dB</th>
<th>ConPR / 24.09dB</th>
</tr>
</thead>
</table>

Fig. 10: Reconstructions on image “Peppers” from $4 \times$ oversampled Fourier magnitude measurements with Poisson noise ($\gamma = 4$).
Fig. 11: Visual reconstruction results of compressive PR on CelebA dataset by different methods at different compression rates.

Fig. 12: Visual comparison of the results of different PR methods from the public dataset.