Variance Stabilization Based Compressive Inversion under Poisson or Poisson-Gaussian Noise with Analytical Bounds

Pakshal Bohra[1], Deepak Garg[2], Karthik S. Gurumoorthy[3] and Ajit Rajwade[2]‡

[1]Department of Electrical Engineering, IIT Bombay
[2]Department of Computer Science and Engineering, IIT Bombay
[3]International Center for Theoretical Sciences, Bengaluru

E-mail:
pakshalbohra@gmail.com,19deepak94@gmail.com,karthik.gurumoorthy@icts.res.in,
ajitvr@cse.iitb.ac.in

1. Proof of Lemmas

Here, we prove the lemmas 3 and 4 from the main paper, which deal with the issue of selection of regularization parameter ρ. The lemmas and their proofs are obtained from Lemma 3.4 of [7].

Lemma 3: $J(\boldsymbol{\theta_\rho})$ is a strictly increasing function of ρ. □

Proof: Consider $\rho_1 < \rho_2$. Let $\boldsymbol{\theta_{\rho_1}}, \boldsymbol{\theta_{\rho_2}}$ be the corresponding minimizers. Then we have

$$\begin{aligned} J(\boldsymbol{\theta_{\rho_1}}) = z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}}) + \rho_1 z_r \|\boldsymbol{\theta_{\rho_1}}\|_1 &\leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}}) + \rho_1 z_r \|\boldsymbol{\theta_{\rho_2}}\|_1 \\ &< z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}}) + \rho_2 z_r \|\boldsymbol{\theta_{\rho_2}}\|_1 = J(\boldsymbol{\theta_{\rho_2}}). \end{aligned} \tag{1}$$

The first inequality follows because $J(\boldsymbol{\theta_{\rho_1}}$ is minimized by $\boldsymbol{\theta_{\rho_1}}$ and the second one is because $\rho_1 < \rho_2$. Thus $J(\boldsymbol{\theta_{\rho_1}}) < J(\boldsymbol{\theta_{\rho_2}})$ when $\rho_1 < \rho_2$. □

Lemma 4: $z_f(\boldsymbol{y}, \boldsymbol{A\theta_\rho})$ is a non-decreasing function of ρ. □

Proof: From Lemma 3, we already have

$$\begin{aligned} z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}}) + \rho_1 z_r \|\boldsymbol{\theta_{\rho_1}}\|_1 &\leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}}) + \rho_1 z_r \|\boldsymbol{\theta_{\rho_2}}\|_1 \\ z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}}) + \rho_2 z_r \|\boldsymbol{\theta_{\rho_2}}\|_1 &\leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}}) + \rho_2 z_r \|\boldsymbol{\theta_{\rho_1}}\|_1. \end{aligned} \tag{2}$$

Dividing the two inequalities by ρ_1 and ρ_2 respectively, we have:

$$\begin{aligned} z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}})/\rho_1 + z_r \|\boldsymbol{\theta_{\rho_1}}\|_1 &\leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}})/\rho_1 + z_r \|\boldsymbol{\theta_{\rho_2}}\|_1 \\ z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}})/\rho_2 + z_r \|\boldsymbol{\theta_{\rho_2}}\|_1 &\leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}})/\rho_2 + z_r \|\boldsymbol{\theta_{\rho_1}}\|_1. \end{aligned} \tag{3}$$

‡ PB and DG are first authors with equal contribution. Corresponding author is AR. AR acknowledges support from IITB Seed Grant 14IRCCSG012. KSG acknowledges the support of the AIRBUS Group Corporate Foundation Chair in Mathematics of Complex Systems established in ICTS-TIFR.

Adding the inequalities, we get:

$$z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}})/\rho_1 + z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}})/\rho_2 \leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}})/\rho_1 + z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}})/\rho_2. \quad (4)$$

Rearranging, we get:

$$(z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}}) - z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}}))/\rho_1 \leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}} - z_f(\boldsymbol{yA\theta_{\rho_2}}))/\rho_2. \quad (5)$$

As $\rho_1 < \rho_2$, it necessarily follows that $z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}}) - z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}}) \leq 0$, i.e. $z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_1}}) \leq z_f(\boldsymbol{y}, \boldsymbol{A\theta_{\rho_2}})$, which means that z_f is a non-decreasing function of ρ. $\square$.

2. Experiments for Poisson CS

We present box-plots for each of the set of results in the main paper for Poisson CS. All the equations are mentioned in the main paper but for completeness we present them here again.

(i) Problem (P3): See Fig. 2.

$$\min \|\boldsymbol{\theta}\|_1 \text{ such that } \|\sqrt{\boldsymbol{y}+c} - \sqrt{\boldsymbol{A\theta}+c}\|_2 \leq \varepsilon, \boldsymbol{\Psi\theta} \succeq \boldsymbol{0}.$$

Here the bound ε was set to $2\sqrt{N}$ based on the tail bound from Theorem 1. Note that the same value of ε was used in all experiments. Problem (P3) was implemented using the well-known CVX package [8] with the SDPT3 solver.

(ii) Problem (P4): See Fig. 1.

$$\min \rho\|\boldsymbol{\theta}\|_1 + \sum_{i=1}^{N}((\boldsymbol{A\theta})_i - y_i \log(\boldsymbol{A\theta})_i), \boldsymbol{\Psi\theta} \succeq \boldsymbol{0}.$$

For (P4), the regularization parameter ρ was chosen omnisciently from the set $\mathcal{S} \triangleq \{10^{-10}, 10^{-9}, ..., 10\}$, i.e. choosing the particular value of $\rho \in \mathcal{S}$ that yielded the least squared difference between the true $\boldsymbol{\theta}$ (assuming it were known) and its estimate. (P4) was implemented using the well-known SPIRAL-TAP algorithm with a penalty for the ℓ_1 norm of DCT coefficients, for a maximum of 500 iterations and with default parameter choices apart from ρ. In all cases, we ensured that the algorithm converged.

(iii) Problem (P5): See Fig. 3.

$$\min \rho\|\boldsymbol{\theta}\|_1 + \|\sqrt{\boldsymbol{y}+c} - \sqrt{\boldsymbol{A\theta}+c}\|_2^2, \boldsymbol{\Psi\theta} \succeq \boldsymbol{0},$$

where ρ was chosen omnisciently from $\mathcal{S}$. (P5) was again implemented using CVX.

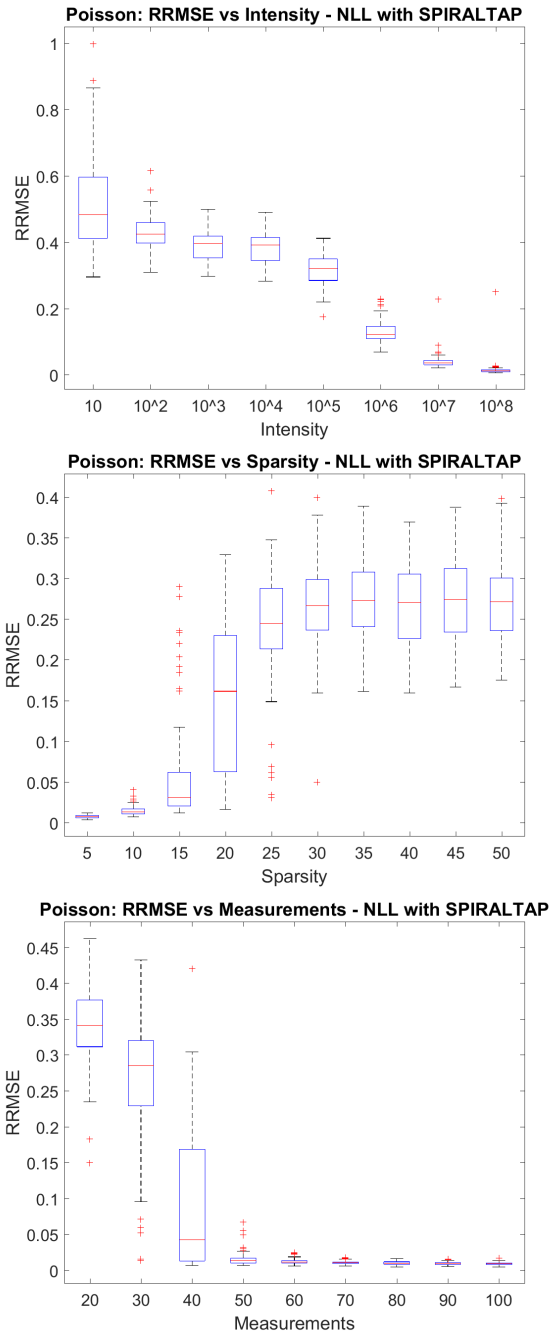

Figure 1. Plots of RMSE($\boldsymbol{x}, \boldsymbol{x}^*$) for (P4) using SPIRAL-TAP with ρ set omnisciently from $\mathcal{S}$. Top to bottom: RRMSE v/s Intensity I at $s = 10, N = 50$; RRMSE v/s Sparsity at $I = 10^8, N = 50$; RRMSE v/s Measurements at $s = 10, I = 10^8$. In each case, reconstruction was for a signal of 100 dimensions.

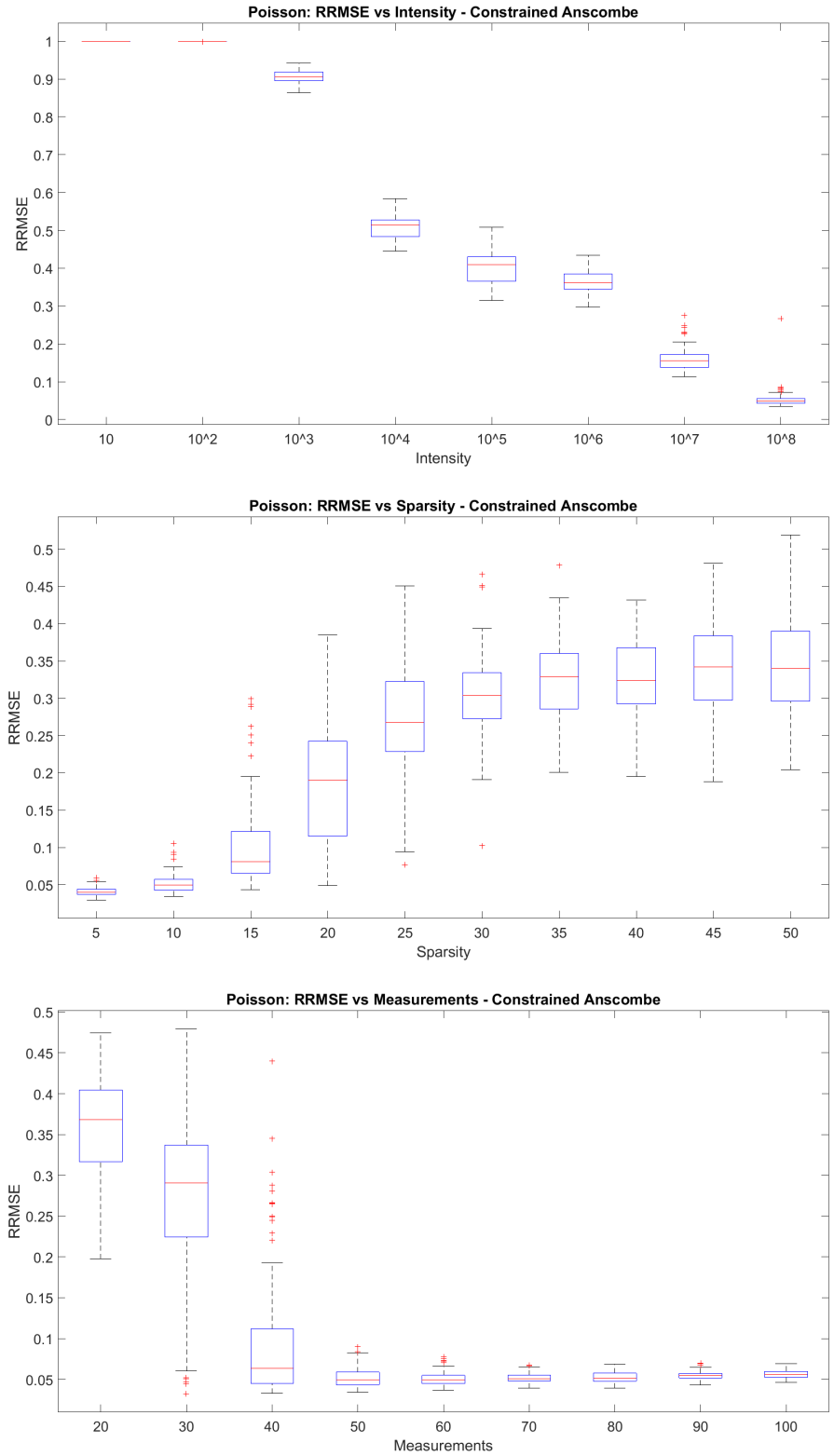

Figure 2. Plots of RMSE($\boldsymbol{x}, \boldsymbol{x}^*$) for (P3) using CVX-SDPT3 with $\varepsilon = 2\sqrt{N}$. Top to bottom: RRMSE v/s Intensity I at $s = 10, N = 50$; RRMSE v/s Sparsity at $I = 10^8, N = 50$; RRMSE v/s Measurements at $s = 10, I = 10^8$. In each case, reconstruction was for a signal of 100 dimensions.

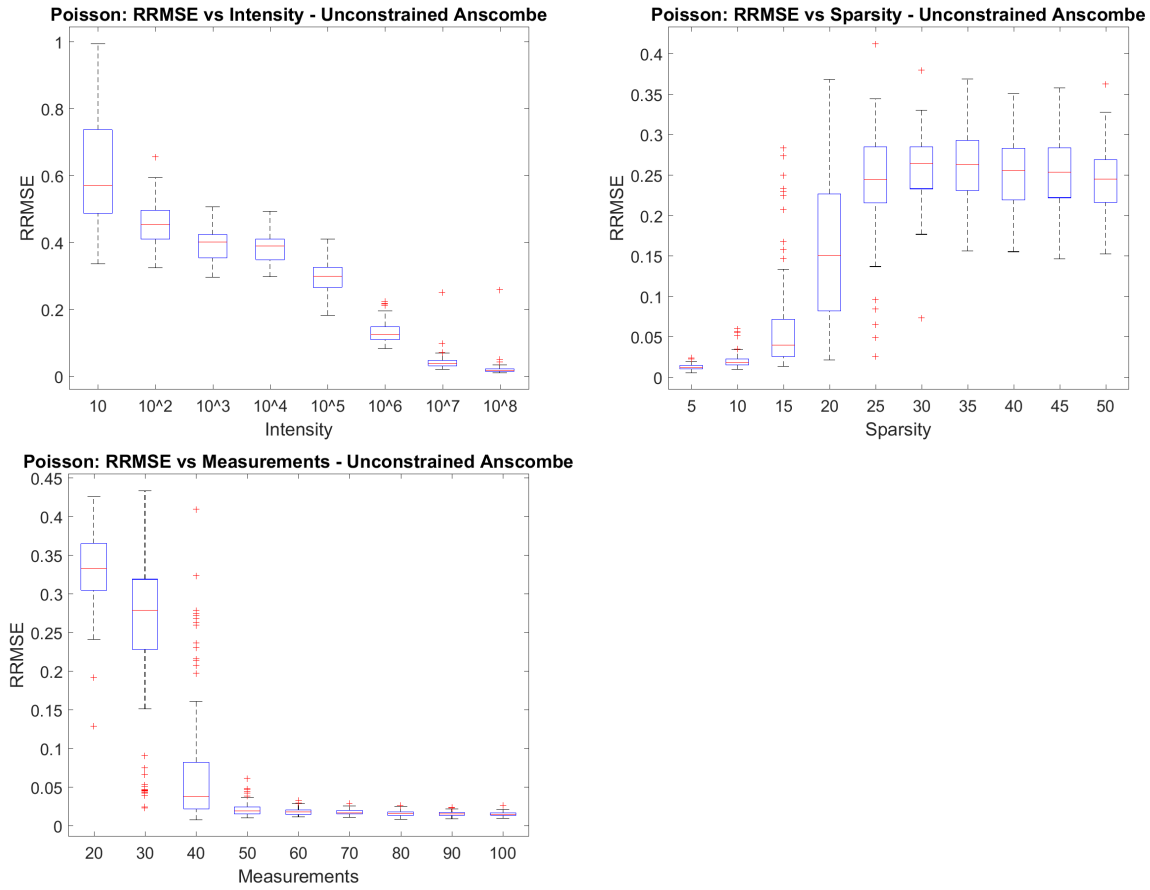

Figure 3. Plots of RMSE$(\boldsymbol{x}, \boldsymbol{x}^*)$ for (P5) using CVX-SDPT3 with ρ set omnisciently from $\mathcal{S}$. Top to bottom: RRMSE v/s Intensity I at $s = 10, N = 50$; RRMSE v/s Sparsity at $I = 10^8, N = 50$; RRMSE v/s Measurements at $s = 10, I = 10^8$. In each case, reconstruction was for a signal of 100 dimensions.

3. Results for Poisson-Gaussian CS

We present box-plots for each of the set of results in the main paper for Poisson CS. All the equations are mentioned in the main paper but for completeness we present them here again.

(i) Problem (PG3): See Fig. 5.

$$(PG3) : \min \|\boldsymbol{\theta}\|_1 \text{ s.t. } \|\sqrt{\boldsymbol{y} + d} - \sqrt{\boldsymbol{A}\boldsymbol{\theta} + d}\|_2 \leq \varepsilon, \boldsymbol{\Psi}\boldsymbol{\theta} \succeq \mathbf{0}, \tag{6}$$

where as defined before $d \triangleq c + \sigma^2$. The bound ε was set to $2\sqrt{N}$. We removed all measurements y_i for which $y_i + d < 0$. Note that the same value of ε was used in all experiments. Problem (PG3) was implemented using the well-known CVX package [8] with the SDPT3 solver.

(ii) Problem (P4): See Fig. 4.

$$\min \rho\|\boldsymbol{\theta}\|_1 + \sum_{i=1}^{N}((\boldsymbol{A\theta})_i - y_i \log(\boldsymbol{A\theta})_i), \boldsymbol{\Psi\theta} \succeq \mathbf{0}.$$

For (P4), the regularization parameter ρ was chosen omnisciently from the set $\mathcal{S} \triangleq \{10^{-10}, 10^{-7}, ..., 10\}$, i.e. choosing the particular value of $\rho \in \mathcal{S}$ that yielded the least squared difference between the true $\boldsymbol{\theta}$ (assuming it were known) and its estimate. (P4) was implemented using the well-known SPIRAL-TAP algorithm with a penalty for the ℓ_1 norm of DCT coefficients, for a maximum of 500 iterations and with default parameter choices apart from ρ. In all cases, we ensured that the algorithm converged.

(iii) Problem (PG5): See Fig. 6.

$$(PG5) : \min \rho\|\boldsymbol{\theta}\|_1 + \|\sqrt{\boldsymbol{y} + d} - \sqrt{\boldsymbol{A\theta} + d}\|_2^2, \boldsymbol{\Psi\theta} \succeq \mathbf{0}.$$

(PG5) was implemented using CVX and using an omniscient choice of $\rho \in \mathcal{S}$ and with removal of measurements for which $y_i + d < 0$.

4. Outlier measurements in Poisson-Gaussian CS

In Poisson-Gaussian CS, measurements for which $y_i + d < 0$ (where $d \triangleq c + \sigma^2$) are considered outlier measurements, because $\sqrt{y_i + d}$ is now complex-valued. We argue here that outlier measurements are rare. Unfortunately, to the best of our knowledge, there are no known tail bounds for Poisson-Gaussian random variables. But by approximating the Poisson-Gaussian distribution for y_i as $\mathcal{N}(\gamma_i, \gamma_i + \sigma^2)$ and applying a Gaussian tail bound, we see that $P(y_i + d \leq 0) \leq \frac{e^{-(\gamma_i + d/\sqrt{\gamma_i + \sigma^2})^2}}{\sqrt{2\pi}[\gamma_i + d/\sqrt{\gamma_i + \sigma^2}]}$. This probability is small, when either γ_i or σ is large. As required by the RIP, our reconstruction bounds hold as long as we have at least $\mathcal{O}(s \log m)$ measurements (out of N) for which $y_i + d \geq 0$. The probability that $y_i + d < 0$ for more than $N - \mathcal{O}(s \log m)$ measurements is even smaller.

References

[1] M. Raginsky, R. Willett, Z. Harmany, and R. Marcia. Compressed sensing performance bounds under Poisson noise. *IEEE TSP*, 58(8):3990–4002, Aug 2010.

[2] X. Jiang, G. Raskutti, and R. Willett. Minimax optimal rates for Poisson inverse problems with physical constraints. *IEEE TIT*, 61(8):4458–4474, 2015.

[3] M.-H. Rohban, V. Saligrama, and D.-M. Vaziri. Minimax optimal sparse signal recovery with Poisson statistics. *IEEE TSP*, 64(13):3495–3508, 2016.

[4] X. Jiang, P. Reynaud-Bouret, V. Rivoirard, L. Sansonnet, and R. Willett. A data-dependent weighted LASSO under Poisson noise. online; accessed July 2016.

[5] Y. Li and G. Raskutti. Minimax optimal convex methods for Poisson inverse problems under lq-ball sparsity. online; accessed July 2016.

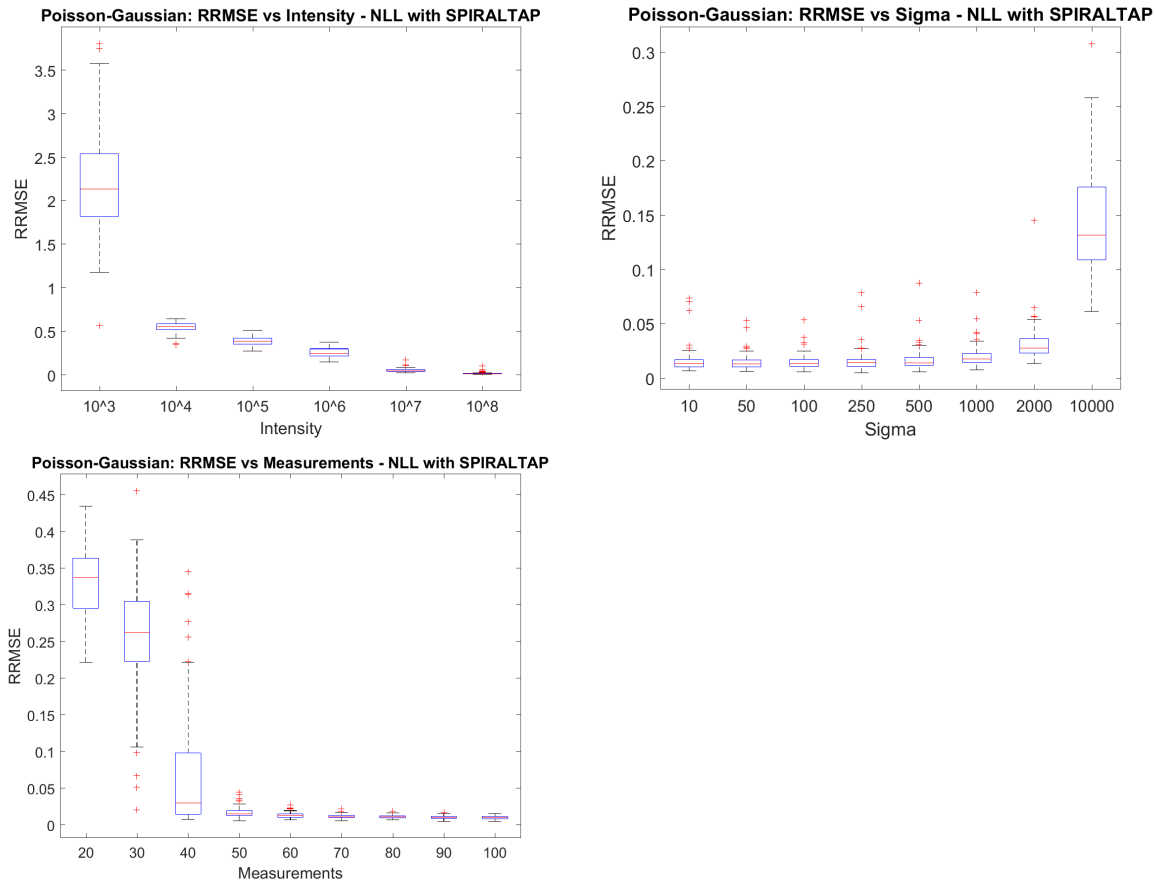

Figure 4. (Plots of RMSE($\boldsymbol{x}, \boldsymbol{x}^*$) for (P4) for Poisson-Gaussian noise using SPIRAL-TAP with ρ set omnisciently from $\mathcal{S}$. Top to bottom: RRMSE v/s Intensity I at $s = 10, N = 50, \sigma = 200$; RRMSE v/s σ at $I = 10^8, N = 50, s = 10$; RRMSE v/s Measurements at $s = 10, I = 10^8, \sigma = 200$. In each case, reconstruction was for a signal of 100 dimensions.

[6] S. Patil, K. Gurumoorthy, and A. Rajwade. Reconstruction error bounds for compressed sensing under Poisson noise using the square root of the Jensen-Shannon divergence. `https://arxiv.org/abs/1606.08557`. Accessed December 2017.

[7] S. Anzengruber and R. Ramlau. Morozovs discrepancy principle for tikhonov-type functionals with nonlinear operators. *Inverse Problems*, 26(2), 2009.

[8] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

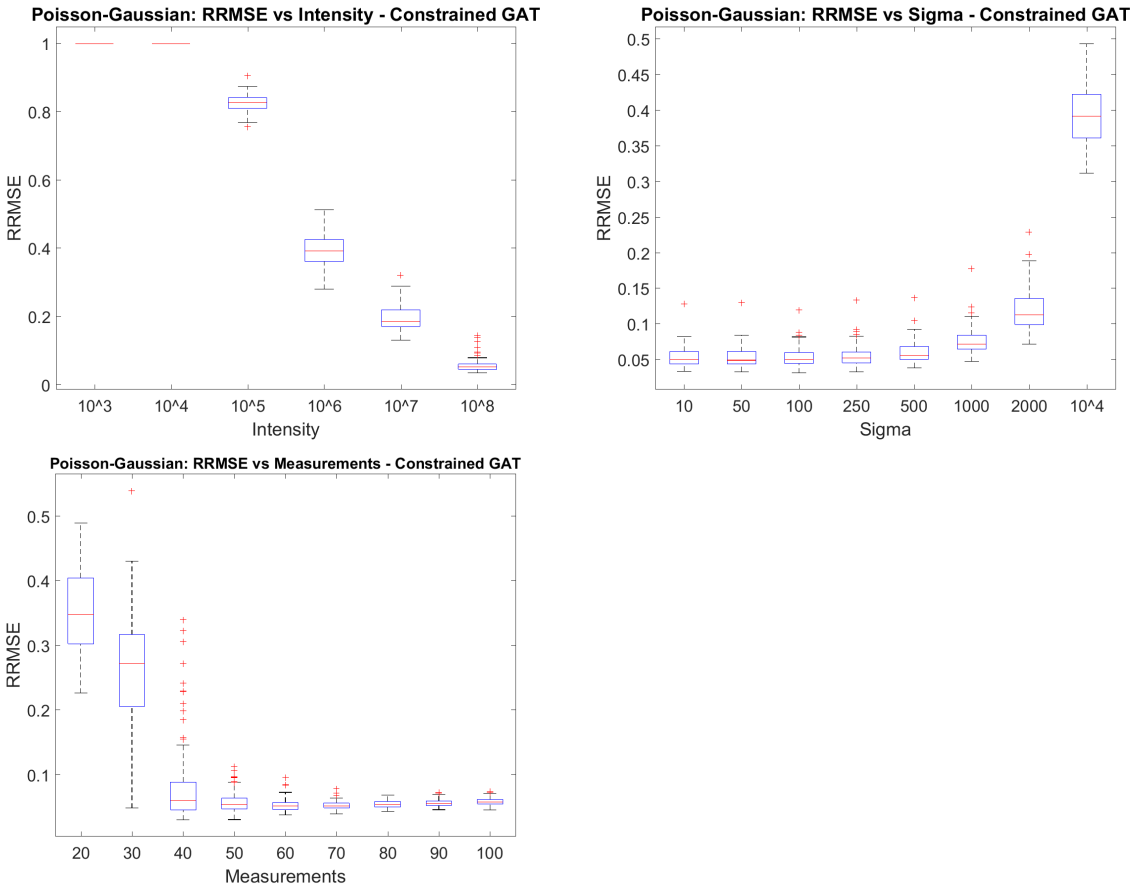

Figure 5. Plots of RMSE($\boldsymbol{x}, \boldsymbol{x}^*$) for (PG3) using CVX-SDPT3 with ρ set omnisciently from $\mathcal{S}$. Top to bottom: RRMSE v/s Intensity I at $s = 10, N = 50, \sigma = 200$; RRMSE v/s σ at $I = 10^8, N = 50, s = 10$; RRMSE v/s Measurements at $s = 10, I = 10^8, \sigma = 200$. In each case, reconstruction was for a signal of 100 dimensions.

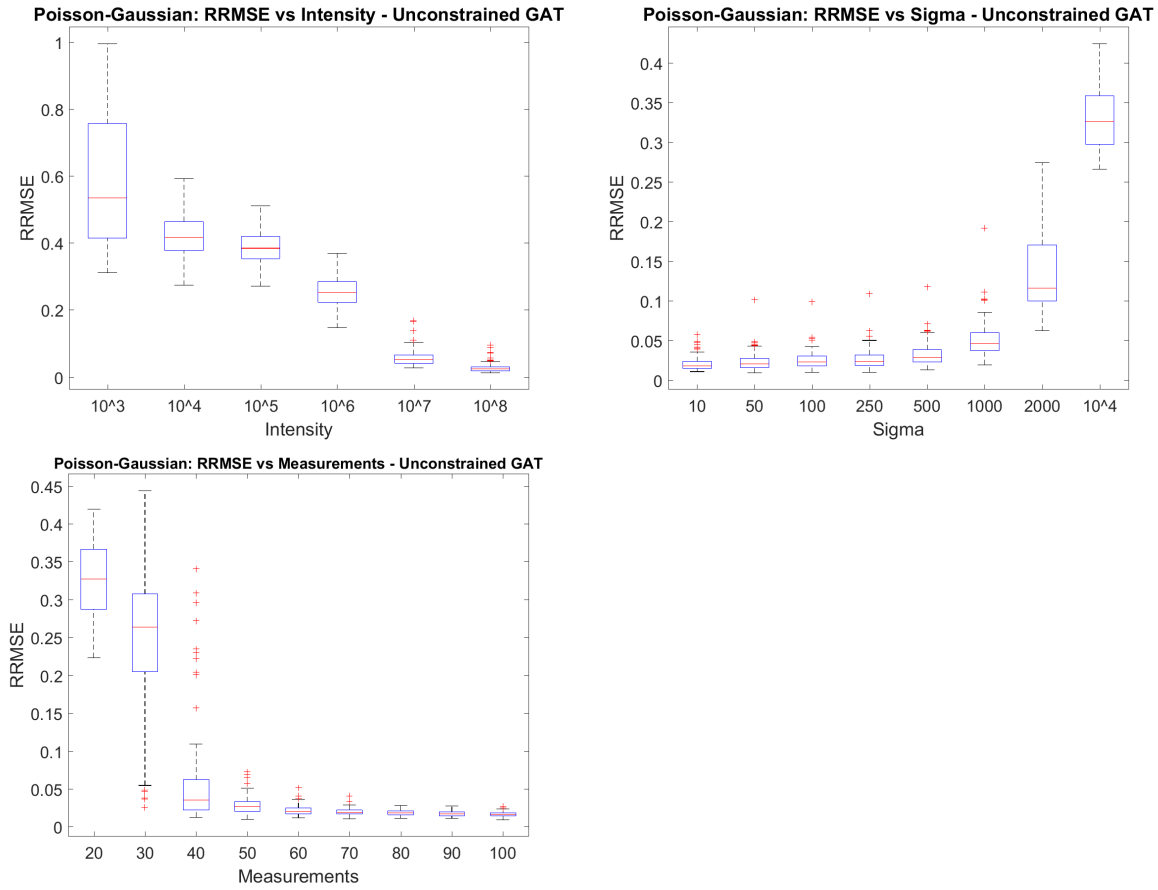

Figure 6. Plots of RMSE($\boldsymbol{x}, \boldsymbol{x}^*$) for (PG5) using CVX-SDPT3 with ρ set omnisciently from $\mathcal{S}$. Top to bottom: RRMSE v/s Intensity I at $s = 10, N = 50, \sigma = 200$; RRMSE v/s σ at $I = 10^8, N = 50, s = 10$; RRMSE v/s Measurements at $s = 10, I = 10^8, \sigma = 200$. In each case, reconstruction was for a signal of 100 dimensions.