An Empirical Assessment of Robotic Sampling Schemes for Environmental Mapping with Compressed Sensing

Susanne Bradley *

Department of Computer Science, University of British Columbia

Abstract

Compressed sensing provides a framework for reconstructing sparse fields using a very small number of measurements. However, ensuring high reconstruction quality involves meeting theoretical requirements that are difficult to satisfy in a robotics context. Some sampling methods have been proposed for compressed sensing-based mapping with mobile robots, but these methods by necessity differ from standard compressed sampling algorithms, and many don't fully meet the prerequisites for good signal reconstruction. In this project we perform simulated experiments to examine the effect this has on signal reconstruction quality, and we find that the performance of established robotic sampling methods is, in practice, similar to that of theoretically well-behaved methods. We find that the robotic methods suffer only slightly, either in terms of a negligible increase ($\sim 2\%$) in the number of samples needed for accurate reconstruction or a moderate increase in the variability of reconstruction quality.

1 Introduction

Environmental mapping is a common function of robots, and many algorithms have been developed for this purpose. However, it is sometimes desirable to do this with as few environmental samples as possible. This might occur if sampling is expensive, or if the domain of interest is large enough that it becomes impractical to sample at every point. The newly invented field of *compressed sensing* (CS) shows that, under certain conditions, it is possible to reconstruct sparse fields using $O(\log n)$ random measurements, where n is the number of sites in the discretized field. Most natural fields are sparse in some basis; if we can sample in such a way that our samples are spread out (or *incoherent*) with respect to that basis, we can employ compressed sensing techniques with a high probability of success.

However, meeting the requirements for good field reconstruction poses unique challenges in a mobile robotics context because of the spatial dependence of robotic trajectories and because robots are restricted in how they sample. As such, many of the sampling methods that have been designed for CS with mobile robots (such as in [10, 16]) do not strictly fulfill the requirements for good CS reconstruction.

A question that has gone largely unaddressed in the literature is whether the lack of theoretical guarantees appreciably impacts reconstruction quality. In this project, I investigate this problem by comparing

*Email: smbrad@cs.ubc.ca

the performance of some established robotic sampling algorithms with that of the theoretically well-behaved Gaussian sampling scheme. In particular, we investigate a random traveling salesperson method, which will meet the compressed sensing requirements for certain problems, and random walk sampling, which, in general, will not. The key contribution of this project is the finding that, in practice, the constraints inherent in robotic sampling effect only small decreases in performance, either in terms of a slight increase in the number of measurements needed to reconstruct a field (for the random traveling salesperson sampling) or increased variability in reconstruction quality (for random walk sampling). These findings suggest that compressed sensing is a promising technique for robotics, despite the existence of some gaps between practical considerations and theoretical requirements.

2 Background and Related Work

2.1 Compressed sensing background

Problem formulation In compressed sensing (CS), we consider the problem of reconstructing a signal from a set of linear measurements, where the number of measurements is considerably less than the dimension of the signal. Mathematically, we have

$$y_{m\times 1} = \Phi_{m\times n} f_{n\times 1}, \tag{1}$$

where we wish to solve for f given a known measurement vector y and sampling matrix Φ. When $m \ll n$, the problem is ill-posed. However, according to compressed sensing, we can perfectly reconstruct the signal using $m = O(\log n)$ measurements under certain assumptions on the matrix Φ and signal f. One of these requirements is that our signal be sparse.

Definition – *sparsity and compressibility*: We say a signal $x \in \mathbb{R}^n$ is *S-sparse* if it has at most S non-zero coefficients. If the remaining $n - S$ coefficients of x are very small rather than strictly zero, we say that x is *compressible*.

A sparse n-element signal has a "true" dimension that is much less than n: the intuition motivating compressed sensing is that we should be able to reconstruct such signals with fewer than n measurements. Our field f may be dense in the spatial domain, but most natural signals are compressible in some basis (Fourier, wavelet, etc.). We can convert f to a compressible signal by computing its representation in an appropriate basis. We define this by:

$$f_{n\times 1} = \Psi_{n\times n} x_{n\times 1}, \tag{2}$$

where x is compressible and Ψ is a basis we will denote as the *sparsity matrix*.

We can now reformulate Equation 1 as

$$y_{m\times 1} = \Phi_{m\times n}\Psi_{n\times n} x_{n\times 1} = \mathbf{A}_{m\times n} x_{n\times 1}, \tag{3}$$

where we call Φ the *sampling matrix* and $\mathbf{A} := \Phi\Psi$ the *observation matrix*.

Signal reconstruction According to Equation 3, our problem is now reduced to solving the underdetermined linear system $y = \mathbf{A}x$ where x is known to be sparse. Ideally, we would like the sparsest solution x that satisfies Equation 3. We can frame this as an optimization problem:

$$\min ||x||_0, \text{ subject to } y = \mathbf{A}x. \tag{4}$$

Unfortunately, this is a non-convex combinatorial optimization problem, so solving it is impractical. However, we can apply convex relaxation [19] to enforce sparsity by minimizing the ℓ_1-norm instead of the ℓ_0-norm, giving us the convex optimization problem

$$\min ||x||_1, \text{ subject to } y = \mathbf{A}x. \tag{5}$$

This reconstruction is exact for S-sparse vectors x (and has low error for compressible vectors) provided that $\mathbf{A}$ satisfies the *Restricted Isometry Property* (RIP) with parameters $(2S, \sqrt{2} - 1)$ [3].

Definition – *Restricted Isometry Property (RIP)* [4]: $\mathbf{A}$ satisfies the RIP with parameters (S, ϵ), $0 < \epsilon < 1$, if

$$(1 - \epsilon)||x||_2^2 \leq ||\mathbf{A}x||_2^2 \leq (1 + \epsilon)||x||_2^2 \tag{6}$$

for all S-sparse vectors x.

In other words, the linear mapping $\mathbf{A}$ roughly preserves the Euclidean length of sparse vectors. An intuitive interpretation is that every set of S columns of $\mathbf{A}$ is close to orthogonal [15].

As of yet, there is no way to construct deterministic matrices that satisfy the RIP. Moreover, checking that a given matrix satisfies the RIP is NP-hard, as you need to check all $\binom{n}{S}$ sets of columns. There are, however, certain classes of random matrices that satisfy the RIP with high probability provided that the number of rows m is greater than $CS\log(\frac{n}{S})$ for some constant C [20]. These include Gaussian, Bernoulli [2], and partial Fourier matrices [19].

2.2 Compressed sensing applications

General Compressed sensing has been used to speed up data acquisition in areas including MRI scans [14], seismology [11], image reconstruction [6], and sensor networks [7], among others. In situations where we can control the entries of the observation matrix $\mathbf{A}$ in Equation 3 directly, most compressed sensing algorithms use independent, identically distributed Gaussian, Bernoulli or uniform entries [5]. As a result, each measurement in our vector y is a dense linear combination of field measurements.

In robotics Compressed sensing was proposed in [10, 16] as a method to increase the efficiency of environmental mapping of large static domains. Potential applications include the use of tethered robots to map atmospheric conditions in forests and using underwater robots to map the chemical and physical properties of water [17]. CS methods have also been used for mapping in [8, 9, 15] as an energy-efficient

alternative to standard SLAM algorithms [12], in which only areas that are directly sensed are mapped. Mostofi et al. [15] developed a CS method for non-invasive obstacle mapping, which allows vehicles to map the inside of a building without entering it. Some of the aforementioned work focuses on purely non-adaptive sampling [10, 16], while others use CS methods as a preprocessing step before performing adaptive sampling on areas of greater interest [8, 9].

The use of compressed sensing for robotic problems presents some unique challenges. The statistical independence of samples required for the RIP assumption is difficult to enforce due to the spatial dependence between a mobile robot's locations over time. The situation can be complicated further by unforeseen obstacles and environmental forces that interfere with movement (e.g. ocean currents). Because a mobile robot has to travel to domain locations to collect samples, taking a dense linear combination of field measurements (as is done in many standard CS algorithms) is impossible.

As a result, we are generally restricted to sampling individual points in the spatial domain, which results in a sampling matrix Φ (Equation 3) that consists of random rows of the $n \times n$ identity matrix (see Figure 1). Assuming that these observations are independent and uniformly distributed over the domain (as is the case with the randomized traveling salesperson algorithm of [10]), then A is likely to satisfy the RIP for some sparsity bases Ψ. A consequence of this is that we can only reconstruct signals that are sparse in certain domains: for example, a signal that is sparse in the Fourier domain is fine because then $\mathbf{A} = \Phi\Psi$ is a random partial Fourier matrix, but we won't be able to reconstruct a signal that is sparse in the spatial domain ($\Psi = \mathbf{I}_n$). With the exception of the random traveling salesperson algorithm, all other robotic sampling methods used in the aformentioned works have spatial dependence between samples, meaning that we have no guarantees that the RIP will hold at all.

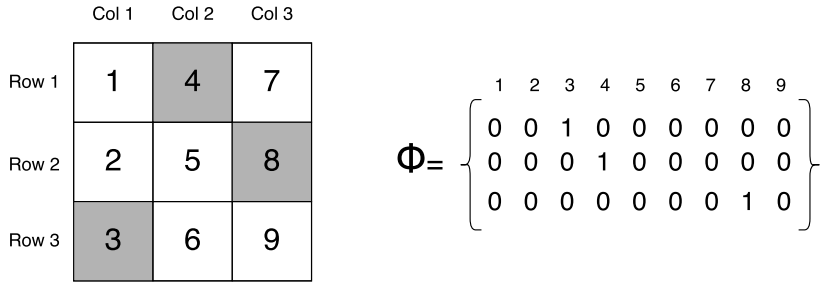
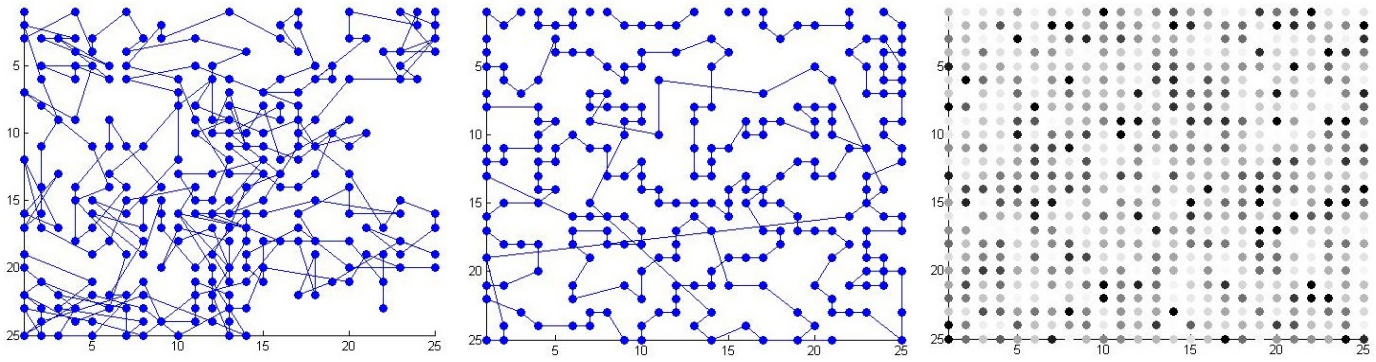


Figure 1: Robotic sampling in the spatial domain. $m = 3$ measurements (in grey) are taken on a 3×3 grid. Each row of the sampling matrix Φ indicates a measurement taken with a 1 in the element corresponding to the location.

3 Sampling Methods

Here we describe the three sampling methods we will use for simulation experiments. Random walk and random traveling salesperson sampling (both introduced in [10]) are designed for mobile robots, and we also perform experiments with Gaussian sampling (one of the preferred methods for compressed sensing algorithms) for comparison. Referring to the factorization of the observation matrix into $\mathbf{A} = \Phi\Psi$, the

purpose of our sampling methods is to generate a suitable sampling matrix Φ. In all our numerical experiments, we will use the discrete cosine transform (DCT) as our sparsity matrix Ψ; we will keep this in mind as we construct our sampling schemes.



(a) (b) (c)

Figure 2: Different sampling schemes on a 25×25 domain. (a) random walk and (b) random TSP with 250 samples. (c) Gaussian sampling: each observation in y is a dense linear combination of field measurements. The figure shows the relative contribution of each field location to a single observation (the darker the point, the more it contributes proportionally to the final observation).

3.1 Random walk sampling

We assume that our robot begins in the top left corner of our domain (position (1,1)). Our tour consists of a series of samples taken after a random step with direction and length drawn from random uniform distributions. The (Euclidean) step length is guaranteed to be at least 1 unit, and the average step length α is a user-specified parameter (i.e. we draw our step length from Uniform$(1, 2\alpha - 1)$). The step direction (in radians) is drawn from a random uniform distribution, the range of which is adjusted according to whether our robot is at an edge or corner of the domain.

If a proposed random step will take us to a location from which we've already sampled, we reject the step and sample again from the length and direction distributions. We perform a maximum of 10 of these rejections before taking a step. This maximum is in place to address the possibility that we're in a part of the space that's been so densely sampled that we're unlikely to reach an unsampled point in a single random step. If a step would take us beyond the edge of the domain, we shorten the step length accordingly.

The key disadvantage of the random walk is that the measurements are pseudo-random and not uniform over the domain – hence, we can't assume that the RIP will hold.

3.2 Random traveling salesperson sampling

The random traveling salesperson (TSP) method was developed in [10] to address the problems of the random walk strategy in fulfilling the RIP. Before the robot begins moving, we choose samples uniformly

at random over the domain and design a TSP tour of the points. As in [10], we use a nearest-neighbours approximation to generate the tour. The points are now truly random instead of quasi-random. In particular, for our numerical experiments where our sparsity basis is the DCT, the observation matrix will satisfy the RIP with high probability.

Despite its theoretical advantages, the random TSP algorithm has some practical disadvantages over the random walk. Since the entire trajectory must be planned in advance, it's difficult to effectively budget the robot's available energy: if we run out of energy before the tour is over, we won't have sampled the domain uniformly; but, if we have extra energy at the end of the tour, there's no obvious way to incorporate additional measurements in a truly random way. The random TSP is also not robust to unforeseen obstacles or other environmental factors that interfere with movement.

3.3 Gaussian sampling

Because the goal of this project is to compare the performance of robotic sampling schemes with theoretically "good" ones, we will compare the reconstruction quality with our previous sampling schemes to that of a Gaussian sampling matrix, which, while being impractical for mobile robots, is a common choice for CS algorithms [5]. We construct $\Phi_{m\times n}$ so that its entries are independently identically distributed (IID) $\mathcal{N}(\mu = 0, \sigma^2 = \frac{1}{m})$, as this is known to satisfy the RIP with high probability [1]. To confirm that Gaussian sampling is a suitable choice for our particular experiments, we need to establish that the observation matrix $\Phi\Psi$ obeys the RIP if Φ is as described and Ψ is the DCT basis.

Proposition 1: Let the sampling matrix $\Phi_{m\times n}$ have IID $\mathcal{N}(0, \sigma^2)$ entries. Then, for any orthonormal $n \times n$ sparsity basis Ψ, the observation matrix $\mathbf{A} = \Phi\Psi$ also has IID $\mathcal{N}(0, \sigma^2)$ entries.

Proof: see Appendix.

Proposition 1 implies that we can use a Gaussian sampling matrix Φ with *any* sparsity basis (as long as it's orthonormal) and be able to reconstruct our signal with high probability. This is the advantage of Gaussian sampling over spatial sampling (such as random TSP), which can only reconstruct signals that are sparse in certain bases.

4 Simulation Experiments

All experiments are performed in MATLAB using the YALMIP toolbox [13] for convex optimization and the MOSEK software package for linear system solves.

4.1 Artificial DCT-sparse datasets

With our first set of experiments, we examine the compressed sensing theorem stating that we can perfectly reconstruct an n-dimensional S-sparse signal using $CS\log(n/S)$ measurements for some constant C [20]. We create random datasets that are sparse in the discrete cosine transform (DCT) domain. We keep S

	Regression results for $m = C \cdot S\log(n/S)$	
Sampling method	C	R^2
Gaussian	1.4019	0.9846
Random walk	1.3296	0.7451
Random TSP	1.4291	0.9078

Table 1: Regression analysis for the number of measurements m needed by the different sampling methods to reconstruct an S-sparse signal as a multiple of $S\log(n/S)$ – the theoretical bound for systems satisfying the RIP.

constant at 50 and use $n = 200, 300, ..., 2000$; we set the length and width of each n-site domain so the domain is as close to square as possible. For each sampling method described in Section 3, we perform 40 trials for each n, using a different random dataset for each of the 40 trials (but using the same datasets for each sampling method). With the random walk, we use an average step length of $\frac{\sqrt{n}}{10}$. For each trial, we compute the minimum number of samples needed for perfect reconstruction, which we define as having a root-mean squared error (RMSE) less than 10^{-10}.

For each sampling method and for each n, we compute the mean and 95% predictive interval of the number of samples needed. Results are shown in Figure 3.

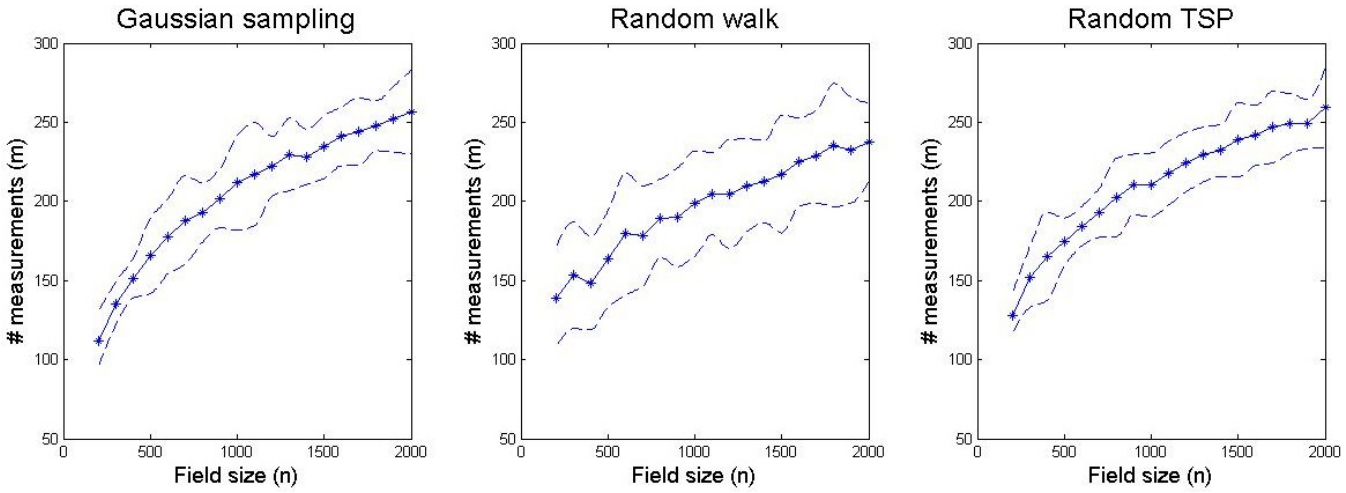


Figure 3: Minimum number of samples needed for each sampling method for random DCT-sparse datasets with different domain sizes and constant sparsity ($S = 50$). For each domain size, we plot the mean number of samples needed for perfect reconstruction (solid lines) as well as the 95% predictive interval for the number of samples needed (dashed lines), based on 40 trials. The predictive interval lines would probably be less noisy if we performed more trials.

We also want to see how well the number of measurements needed for perfect reconstruction with each sampling method fits with the theoretically predicted value of $CS\log(n/S)$. To this end, we fit a linear regression model describing the mean number of measurements needed as a multiple of $S\log(n/S)$. The least-squares estimate of C along with the coefficient of determination R^2 for each method is given in Table 1.

Discussion First, we note from Figure 3 that the random TSP and Gaussian sampling perform similarly in terms of both mean and variance of the number of measurements needed, with Table 1 showing that the Gaussian approach generally requires about 2% fewer measurements. Somewhat surprisingly, the random walk outperforms the other methods on average, but has about twice their variance (hence its wider confidence bands in Figure 3). Also note from the R^2 values in Table 1 that the random walk is the one method where the number of measurements needed is not well-predicted by an $O\log(n/S)$ formula (which is unsurprising, as it is the only sampling method that does not satisfy the RIP). For robotic sampling, it's clear that you will, on average, get satisfactory results with a random walk, but the random TSP approach is "safer": that is, its variance is lower and the number of measurements needed for signal reconstruction tends to agree with the bounds given by compressed sensing theory.

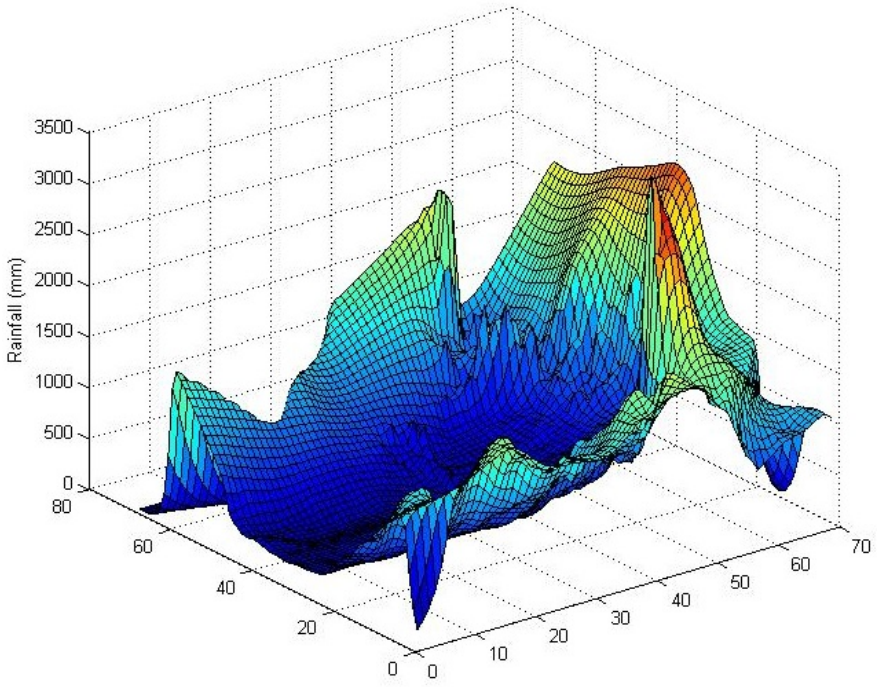
4.2 Environmental dataset

We now compare the reconstruction error of our sampling methods on a real dataset consisting of average annual rainfall in Australia between 1961 and 1990 [18]. The original dataset consists of the mapped data (Figure 4b) projected onto a 690×880 grid. A dataset this size would result in sampling and sparsity matrices too large for the implementation of MATLAB I was using, so we crop out the western portion of the map and coarsen the grid so that its size is 69×69 (this is the same size as the environmental field that was reconstructed in [16]). This signal (shown in Figure 4a) is compressible in the DCT domain: 99.5% of the signal's energy is captured by 5% of the DCT coefficients.

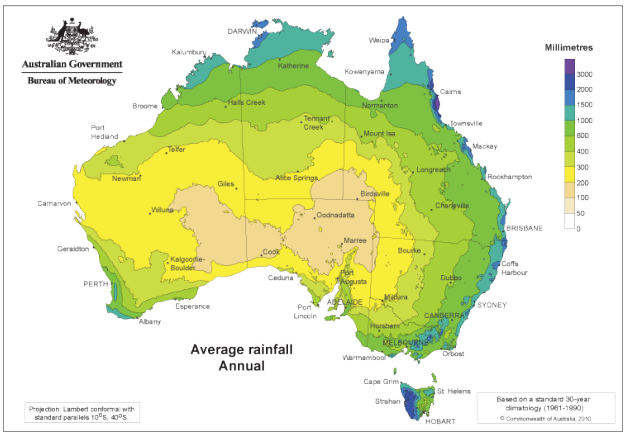
For each sampling method, we attempt in a single trial to reconstruct the field with different numbers of samples ranging between 100 (2% of the field) and 2400 (50.4% of the field). For the random walk, we use an average step length of 6 units. In each case, we measure the normalized RMSE, given by dividing the RMSE by the range of the field ($f_{\max} - f_{\min}$). For this experiment, we find that all methods give similar performance but random walk again suffers from high variability – note the occasional large jumps in reconstruction error in Figure 5. For robotic sampling of a real dataset, it appears that random TSP is a better sampling method. Excepting cases where we have very few measurements (≤ 200), its performance is virtually indistinguishable from that of the Gaussian sampling. Field reconstruction using random TSP sampling with 1000 and 2300 samples is shown in Figure 6.

5 Conclusions and Future Work

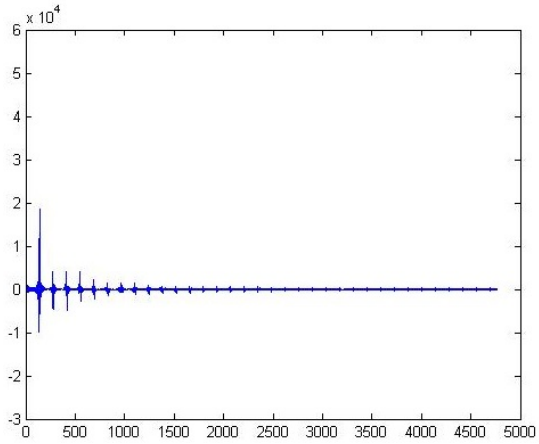
We compared sampling schemes that are feasible for mobile robots with Gaussian sampling, which is a method of choice for standard compressed sensing algorithms. For our robotic sampling methods, we examined the random TSP, which obeys the RIP with high probability for certain sparsity bases including the discrete cosine transform; and the random walk, about which we have no such guarantees. We find that the robotic sampling schemes perform almost as well as the theoretically "good" Gaussian sampling method. In particular, the performance of the random TSP is very similar to that of the Gaussian sampling,



(a)



(b)



(c)

Figure 4: We reconstruct data of mean annual rainfall in Australia ((b) – image courtesy of [20]). The data is available from [20] as a projection onto a 2D rectangular grid, which we coarsen to a manageable size (69×69) for signal reconstruction (a). This signal is compressible in the DCT domain – plot of the DCT coefficients is shown in (c).

and its performance is well-predicted by compressed sensing theory, making it a good choice for reliable reconstruction quality. Despite a lack of theoretical performance guarantees, the random walk gives high reconstruction quality in practice, though its performance is more variable than that of the two RIP-compliant schemes.

A limitation of this work is that the small numbers of trials used in both types of experiments have led

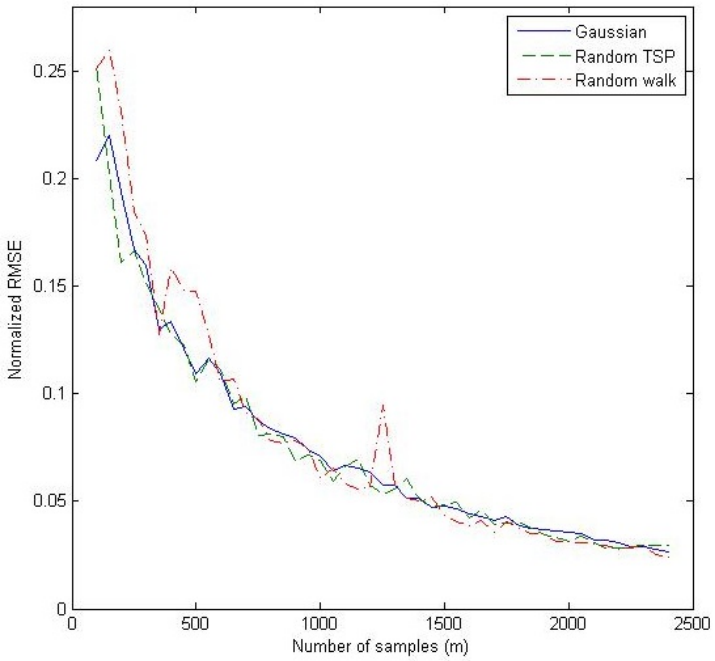
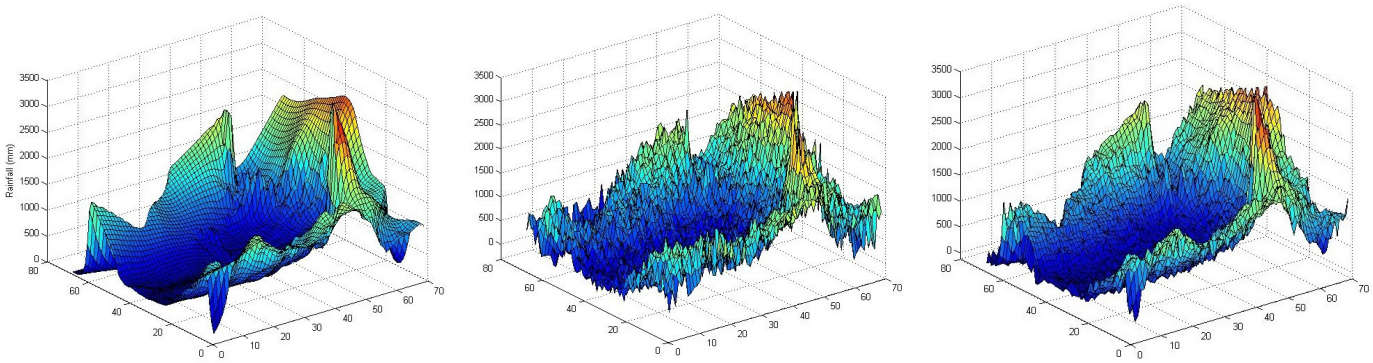


Figure 5: Normalized RMSEs for a single trial of reconstructing the map of Australian rainfall data with varying numbers of samples.



(a) (b) (c)

Figure 6: Reconstruction of a 4761-element map of rainfall data (a) using random TSP with 1000 samples (b) and 2300 samples (c). Normalized RMSEs are 0.072 and 0.027, respectively.

to somewhat high-variance estimates of the effectiveness of our sampling. In addition, because of limited computational resources, we were unable to perform field reconstruction for datasets of a size where we would realistically want to perform compressed sensing. Specifically, I only had sufficient memory to perform the necessary optimization for fields of dimension under 5000. In addition, this work does not consider the effects of complicating factors such as obstacles, environmental disturbances, or imprecise localization that often occur in real robotic mapping problems.

A possibility for future work in this area would be a theoretical analysis of the behaviour of random

walk-type sampling – or, more generally, sampling with spatial dependence – as this could allow for more flexibility in performing compressed sensing with mobile robots. In addition, it remains to be explored how these methods would scale in practice to very large domains.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[2] E. Candes. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. des Sci Serie I*, 346:589–592, 2008.

[3] E. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, Dec 2006.

[4] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[5] D. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, April 2006.

[6] K. Egiazarian, A. Foi, and V. Katkovnik. Compressed sensing image reconstruction via recursive spatially adaptive filtering. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 1, pages I – 549–I – 552, Sept 2007.

[7] F. Fazel, M. Fazel, and M. Stojanovic. Random access compressed sensing for energy-efficient underwater sensor networks. *Selected Areas in Communications, IEEE Journal on*, 29(8):1660–1670, September 2011.

[8] M. Horning, M. Lin, S. Srinivasan, S. Zou, M. Haberland, K. Yin, and A. Bertozzi. Compressed sensing environmental mapping by an autonomous robot, 2015.

[9] S. Huang and J. Tan. Compressive mobile sensing in robotic mapping. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS'09, pages 3070–3075, Piscataway, NJ, USA, 2009. IEEE Press.

[10] R. Hummel, S. Poduri, F. Hover, U. Mitra, and G. Sukhatme. Mission design for compressive sensing with mobile robots. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2362–2367, May 2011.

[11] F. J.Herrmann, M. P.Friedlander, and O. Yilmaz. Fighting the curse of dimensionality: compressive sensing in exploration seismology. *IEEE SIgnal Processing Magazine*, 29:88–100, May 2012.

[12] J. Leonard and H. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on*, pages 1442–1447 vol.3, Nov 1991.

[13] J. Lofberg. Automatic robust convex programming. *Optimization methods and software*, 27(1):115–129, 2012.

[14] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[15] Y. Mostofi. Compressive cooperative sensing and mapping in mobile networks. *Mobile Computing, IEEE Transactions on*, 10(12):1769–1784, Dec 2011.

[16] S. Poduri, G. Marcotte, and G. S. Sukhatme. Compressive sensing based lightweight sampling for large robot groups, 2011.

[17] R. Pon, M. A. Batalin, J. Gordon, A. Kansal, D. Liu, M. Rahimi, L. Shirachi, Y. Yu, M. Hansen, W. J. Kaiser, M. Srivastava, G. Sukhatme, and D. Estrin. Networked infomechanical systems: A mobile embedded networked sensor platform. In *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, IPSN '05, Piscataway, NJ, USA, 2005. IEEE Press.

[18] Australian Government Bureau of Meteorology. Average annual, seasonal and monthly rainfall, 2015. Online, `http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.jsp`: accessed 2015-12-12.

[19] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and gaussian measurements. In *Information Sciences and Systems, 2006 40th Annual Conference on*, pages 207–212, March 2006.

[20] X. Shen and Y. Gu. Restricted isometry property of subspace projection matrix under random compression. *CoRR*, abs/1502.02245, 2015.

Appendix.

Proof of Proposition 1: Let $Z_1, Z_2, ..., Z_m$ denote the *rows* of Φ and $\psi_1, \psi_2, ..., \psi_n$ the *columns* of Ψ:

$$\Phi = \left(\begin{array}{c} Z_1 \\ \hline Z_2 \\ \hline \vdots \\ \hline Z_m \end{array} \right), \Psi = \left(\begin{array}{c|c|c|c} & & & \\ \psi_1 & \psi_2 & ..... & \psi_n \\ & & & \end{array} \right).$$

Each $Z_i = (z_{i,1}, z_{i,2}, ..., z_{i,n})$ consists of independent $\mathcal{N}(0, \sigma^2)$ entries. The joint density of $z_{i,1}, z_{i,2}, ..., z_{i,n}$ is

$$\begin{aligned} f(z_{i,1}, z_{i,2}, ..., z_{i,n}) &= f(z_{i,1}) \cdot f(z_{i,2}) \cdot ... \cdot f(z_{i,n}) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{z_{i,1}^2 + z_{i,2}^2 + ...z_{i,n}^2}{2\sigma^2} \right) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{||Z_i||_2^2}{2\sigma^2} \right). \end{aligned}$$

Because this distribution depends only on the length of Z_i, we say it is *rotationally invariant*, i.e. the joint distribution remains the same under rotations in n-dimensional space. Now, we know that $\langle Z_i, \mathbf{e}_1 \rangle = z_{i,1}$ has an $\mathcal{N}(0, \sigma^2)$ distribution. But, because of the rotational invariance, $\langle Z_i, u \rangle$ has the same $\mathcal{N}(0, \sigma^2)$ distribution for any unit vector u.

Therefore, $\mathbf{A}_{i,j} = \langle Z_i, \psi_j \rangle \sim \mathcal{N}(0, \sigma^2)$ for all $1 \leq i \leq m, 1 \leq j \leq n$.

We have now proven that the entries of $\mathbf{A}$ have the same distribution as those of Φ. We still need to prove that these entries are independent. This is clearly the case for entries in different rows, as these are linear transformations of independent vectors Z_i.

Now, consider two entries in the same row: $X := \mathbf{A}_{i,j} = \langle Z_i, \psi_j \rangle$ and $Y := \mathbf{A}_{i,k} = \langle Z_i, \psi_k \rangle$, where $j \neq k$. Because X and Y have a joint multivariate normal distribution, they are independent if and only if they are uncorrelated. Therefore, we need only check that $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$. Because $\mathbb{E}(z_{i,j}) = 0$ for all i, j, we have:

$$\begin{aligned} \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) &= \mathbb{E}(XY) \\ &= \mathbb{E}[(z_{i,1}\psi_{1,j} + z_{i,2}\psi_{2,j} + ... + z_{i,n}\psi_{n,j})(z_{i,1}\psi_{1,k} + z_{i,2}\psi_{2,k} + ... + z_{i,n}\psi_{n,k})] \end{aligned}$$

Because the z's are independent and mean-zero, the cross terms cancel, leaving:

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}[\psi_{1,j}\psi_{1,k}z_{i,1}^2 + \psi_{2,j}\psi_{2,k}z_{i,2}^2 + ... + \psi_{n,j}\psi_{n,k}z_{i,n}^2] \\ &= \psi_{1,j}\psi_{1,k}\mathbb{E}(z_{i,1}^2) + \psi_{2,j}\psi_{2,k}\mathbb{E}(z_{i,2}^2) + \psi_{n,j}\psi_{n,k}\mathbb{E}(z_{i,n}^2) \\ &= \sigma^2\langle \psi_j, \psi_k \rangle \\ &= 0, \text{ by orthogonality of } \Psi. \ \square \end{aligned}$$