One-step-ahead kinematic compressive sensing

Citation: Hover, Franz S. et al. "One-step-ahead Kinematic Compressive Sensing." Proceedings of the 2011 IEEE GLOBECOM Workshops (GC Wkshps), 2011. 1314–1319.

As Published: http://dx.doi.org/10.1109/GLOCOMW.2011.6162400

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/78584

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

One-Step-Ahead Kinematic Compressive Sensing

F.S. Hover and R. Hummel
Dept. Mechanical Engineering
Massachusetts Institute of Technology
Cambridge MA 02139
Email: hover@mit.edu, hummel@mit.edu

U. Mitra and G. Sukhatme
Departments of Electrical Engineering
and Computer Science
University of Southern California
Los Angeles CA 90089-2905
Email: ubli@usc.edu, gaurav@usc.edu

***Abstract*—A large portion of work on compressive sampling and sensing has focused on reconstructions from a given measurement set. When the individual samples are expensive and optional, as is the case with autonomous agents operating in a physical domain and under specific energy limits, the CS problem takes on a new aspect because the projection is column-sparse, and the number of samples is not necessarily large. As a result, random sampling may no longer be the best tactic. The underlying incoherence properties in l_0 reconstruction, however, can still motivate the purposeful design of samples in planning for CS with one or more agents; we develop here a greedy and computationally tractable sampling rule that will improve errors relative to random points. Several example cases illustrate that the approach is effective and robust.**

I. Introduction

Compressive sampling and its related methodologies [1],[2] have enjoyed great success in varied domains including geophysics, synthetic aperture radar, medical imaging, and many others. A key result is that, subject to signal sparsity and uncorrelated measurements, exact reconstruction of a signal is possible with an unusually small number of randomly-chosen samples; the critical number scales with the sparsity and the log of the number of degrees of freedom. Variations for signals that are not strictly sparse, and for scenarios with sensor noise, have been developed also, subject to the restricted isometry property.

Because CS reconstructs signals with few measurements, it is a reasonable question as to how CS could be employed when each individual sample has a specific cost. Indeed, if spatial sampling is to be carried out by a mobile agent moving through a domain, operational issues arise immediately that give the CS problem a different flavor. A nonzero cost of transit between samples would argue for a traveling salesman (or similarly efficient) path visiting a fixed set of random points, but there is also a practical desire to be robust against premature termination of the CS mission. This argues instead for suitably-sized random steps taken incrementally across the space. Conventional regular paths, such as the cellular boustrophedon in two-space [3] are analogous to the TSP in the sense that premature termination is likely to destroy any chances of a meaningful reconstruction from the data in hand. We note that although samples can often be taken in transit, from the CS point of view these are by definition spatially correlated with the endpoints, and therefore in some sense redundant and less valuable than the vertices.

A compelling application today is the deployment of autonomous vehicles in the ocean, that sample slowly-changing water properties over a very large domain. The ability to perform strong reconstruction from a small number of well-chosen sample locations would improve our understanding of important ocean features such as oil plumes and other structures [4], [5]. Increasingly, multiple agents are being considered for such applications, and underwater they are usually connected by wireless acoustic networks. As we will discuss, the use of CS in designing missions is complementary with wireless networking.

Our goal in this paper is to describe and develop the basic kinematic compressive sensing problem. In the following section we state it more fully and discuss how operational constraints affect compressive sensing. We propose a simple greedy algorithm - called RIPD - for ranking sample locations in a way that balances computational and operational constraints with reconstruction error. Modest but consistent improvements are demonstrated in the single-vehicle case with discrete cosine and Haar wavelet transforms.

II. Problem Scenario

In the canonical case of CS, an under-determined system $Ax = z$ is solved while minimizing $||x||_0$, in many cases equivalent to minimizing $||x||_1$, for which fast algorithms exist. Here A is the effective dictionary, x the vector of modal coefficients for the dictionary, and z the set of m CS measurements. Denoting N as the number of degrees of freedom in the system, if x is known to have $S < N$ components that are nonzero, the signal is said to be strictly S-sparse, and exact reconstruction is possible with a small number of incoherent measurements.

CS methods carry the general recommendation of random samples, for the purpose of reducing their coherence, in probability, with respect to the reconstruction basis. We can write $A = PD$, where D is the underlying, possibly redundant dictionary matrix and P is a projection matrix that is randomly chosen, or could be designed. We thus differentiate between $z = Py$, the CS measurements, and $y = Dx$, the set of physical measurements that are possible. The physical measurements are the complete set of pixel values in an image, for instance, whereas the CS measurements are a smaller set of their linear combinations, as encoded in P. Thus if D is complete (square), then P is $m \times N$, with $N \gg m$.

We are aware of a few works treating the design of optimal projection matrices for a given dictionary. Elad [6] addressed the problem by minimizing mutual coherence of the effective dictionary A. The method iteratively reduces large elements in the Gram matrix $A^T A$ directly, and then computes a new projection satisfying rank conditions associated with the desired number of measurements. Duarte-Carvajalino and Sapiro [7] addressed the same objective shortly thereafter as an eigenstructure problem, and achieved faster computations; Xu *et al.* [8] followed with an approach based on the equiangular tight frame, and reported further improvements, although computing times are not given. These methods put no constraints on P and the outcome of optimization is dense in the general case, because implicit is the assumption that all of y is available. We also note that very specific constraints in measurement matrix design have been developed in the MIMO radar community; here the task is to choose transmission waveforms and array parameters so as to create incoherence in the sensing matrix PD [9],[10],[11].

When some elements of y are not used or available, the projection P has $N - m$ zero columns. This is a significant specification, because in contrast with the unconstrained versions above, the problem is now combinatorial and likely to be exponential in cost. Column-sparsity of P is indeed the major consequence when point data is collected by a mobile agent in a physical domain. Additional factors can come into play: motions of the agent may be subject to speed, attitude rate, energy, communication, and collision constraints. If the vehicle moves too slowly, the field may change. Thus we see that the collection of data for compressive sensing is a rich planning problem, and the insertion of transit costs into CS is only one of many possible aspects.

We first addressed kinematic CS in [13]. In comparing sparse reconstruction from points obtained through a heuristically-optimized random walk, through a TSP over random points, and through a quasi-random boustrophedon strategy, we found no essential differences: assuming that the mission length allows multiple traversals of the space, the leading factor in reconstruction performance is simply the number of points. A key question then becomes the planning horizon, because both random-point TSP and randomized lawnmowing approaches are efficient only with long horizons, whereas a random walk has no particular horizon, beyond the implicit assumption that the agent crosses the space adequately. More specifically, we distinguish among three main paradigms for collection of data:

- A suitably long planning horizon such that $m \geq 4S$ is guaranteed. If a full set of points is chosen and they are not located trivially, they can be visited using a traveling salesman path, whose distance is approximately $0.75\sqrt{m}$ in the unit square [14].
- A one-step ahead horizon, to generate the best CS product possible from points sampled so far, plus one additional.
  - Transit is cheap so step lengths are unconstrained.
  - Transit is expensive and so it is desired to collect samples frequently - this is accomplished by limiting the transit length.

A significant factor also is that the number of collected samples may not be large in kinematic CS, numbering perhaps in the hundreds or even tens for certain applications. This calls into question the suitability of fully random point selections, which are justified only in probability.

An important and broad extension to the basic problem involves multiple agents communicating wirelessly; we make a few initial observations. First, a network enables centralized real-time planning, as well as a centralized reconstruction process that can continuously perform reconstructions and thereby make decisions, for example, on whether enough points have been obtained. Second, designed CS sample locations can be visited by partitioned TSP paths [12] that are efficient if the horizon is long; these would require long-range connectivity. A one-step-ahead policy can also accommodate physical partitions explicitly, but might instead improve connectivity by moving the vehicles as an unstructured fleet and avoiding collisions. Finally, as described we are focusing on point measurements in a physical space; but multiple networked vehicles are also well-suited to tomographic measurements taken across the domain (as in MRI scans), in which case the designed CS samples are specific cross-sections.

We present and demonstrate below a simple greedy method for actively minimizing coherence. It is developed in the context of a single vehicle, but the concept applies also to networked groups of vehicles. As we show in examples, reconstruction errors appear to be systematically improved over random point selection, but two caveats are noted. The method incurs a significant computation load, although it is well within the capability of modern microprocessors during transit time for undersea and many other types of mobile agents. Second, the method requires that the reconstruction basis be specified at the time of sampling, although it can be changed on the fly with an added computation. We will not consider in this method the learning of dictionaries, or non-square dictionaries.

III. Formulation

The specific condition for exact reconstruction in large-scale CS is that $m \geq C\mu^2(\phi, \psi) S \log(N)$[1]. Here, C is a constant, and $\mu(\phi, \psi)$ captures the incoherence of the reconstruction basis ϕ and the measurement basis ψ. As noted above, most compressive sensing work to date advocates choosing samples randomly, because this almost surely minimizes μ. Related to incoherence, the bases are said to satisfy the restricted isometry property if each column of the effective dictionary A is nearly orthogonal to every other column. A scenario satisfying RIP will be amenable to reconstruction with noise and when the signal is not strictly sparse, but still compressible.

For the purpose of describing our construction, let us say that m measurements have been taken already. Hence there is an $m \times N$ A-matrix, which we denote A_m, so that $A_{m,i}$ is the i'th column of A_m. The last measurement is z_m, taken at location r_m. The decision is where to take the next

measurement. The best we can do in a greedy algorithm is to choose the measurement location r_{m+1} so as to minimize the resultant coherence of the whole set, which is

$$q(A_m, r_{m+1}) = \max_{i,j} |\langle A_{m+1,i}, A_{m+1,j} \rangle|.$$

Here $r_{m+1} \in R(p, r_1, r_2, \ldots, r_m)$, the feasible set of additions given all the previous measurement locations, and parameters p. These encode for example the constraints of no repeated points, limited stepsize, and domain boundaries, as well as others. For any non-trivial feasible region defined by R, however, the function q is complicated, with multiple local minima and discontinuities depending on the transform in use; this leads to a sampling-based strategy. The idea is to select t trial locations $r_k^- \in R, 1 \leq k \leq t$, and set $r_{m+1} = r_k^-$ minimizing $q(A_m, r_k^-)$. This is the greedy method we are proposing in Algorithm 1, referred to as RIP Design, or RIPD. The algorithm adds a points to a base set of m points. Note that in our examples below we use m to denote the total number of samples, consistent with the literature.

The procedure of assessing candidates and choosing one at each cycle is constant-time and has cost that scales with tN^2. To see this, let Q be the symmetric $N \times N$ matrix whose diagonal is zero and whose off-diagonal terms are the inner products induced by A_m. This is the Gram matrix, and coherence q is the maximum absolute value among elements in Q. Assessing Q^- for a candidate addition r^- is simple:

$$Q^- = Q + A(r^-)^T A(r^-),$$

where $A(r^-)$ is the projection onto the reconstruction basis; $A(r^-)$ is the row that would be added onto A_m were the point r^- accepted.

The ensemble of t candidates can be chosen by any method desired. For instance, if it is desired to cross the domain quickly, then one would put them on the boundaries of R, either deterministically or randomly. They can be placed uniformly random in the interior of R, as in our examples below. In the case of obstacles, feasible paths can be determined using a fast, and (typically) sub-optimal scheme such as the RRT [15]. Herein, we see that a new optimization problem arises that would weigh the long-term statistical improvement in reconstruction error (via greedily reducing coherence) against the very specific costs and risks of a complicated path through obstacles. This is outside our scope at present.

We make a few implementation notes and observations. As indicated in the algorithm listing, the Gram matrix diagonal should not be used for the purpose of selecting from the candidate points. Changing basis on the fly presents no difficulty, with the understanding that A_m and Q have to be recomputed each time it is changed. The greedy technique approaches a deterministic sequence as t grows, because the candidate points will identify the global minimum for q. Such design locations would depend only on the initial locations and therefore become correlated, but it is important to note that it is not in the reconstruction basis. As a result, increasing t as computational resources permit will tend to improve the reconstruction performance.

Algorithm 1 $r = RIPD(A_m, t, a, r_1, r_2, \cdots, r_m, p)$

$Q \leftarrow A_m^T A_m$
for $n = 1 : a$ **do**
  for $k = 1 : t$ **do**
    $r_k^- \leftarrow \text{sample}(R(p, r_1, r_2, \cdots, r_{m+n-1}))$
    $q_k \leftarrow \max |Q + A(r_k^-)^T A(r_k^-)|$ (element-wise max, excluding terms on the diagonal)
  end for
  $l \leftarrow \text{argmin } q$
  $r_{m+n} = r_l^-$
  $Q \leftarrow Q + A(r_l^-)^T A(r_l^-)$
  $A_{m+n} \leftarrow [A_{m+n-1}; A(r_l^-)]$
end for

IV. Examples

We now address several example computations; these are meant to show typical behavior of our greedy, RIP-based design, but not to systematically characterize its behavior. In the first we consider the cosine basis, which is supported across the continuous measurement domain, and the effects of sensor noise estimation errors. In the second we study the Haar wavelets, involving unsupported basis functions on a discrete domain; there is no sensor noise in this case, but we vary the allowable transit distance. Our statistics are taken over five hundred trials, and our standard for comparison is a reconstruction based on randomly chosen points, termed RP.

A. Discrete-Cosine Transform in One and Two Dimensions with Noise

We use the discrete-cosine transform (DCT) in our first two cases, over a unit domain; the space is considered continuous. The specific problem parameters are $S = 15$ (sparsity), $N = 196$ (degrees of freedom), and $t = 100$ (candidates). We sample one point randomly and then proceed to design another forty-nine, so that $m = 10S/3$, a slightly lower value than is typical for exact reconstruction. We apply normally distributed, zero-mean noise to all measurements, with standard deviation ν equal to one percent of the RMS signal level, and solve the modified problem with weighted l_1 minimization as given by [16]: $\min_x ||x||_1$ s.t. $||Ax - b||_2 < \hat{\nu}\sqrt{m + \sqrt{2m}}$. We do not include kinematic constraints in the one-dimensional case.

As seen in Figure 1, the random and RIPD-selected points are indistinguishable viewed directly. Figure 2 shows the normalized reconstruction errors as a function of estimated noise level divided by true noise level; the zero log value on the horizontal axis corresponds to exact matching. Normalized reconstruction error is taken as the Euclidean norm of the error in y divided by the norm of the signal variation from its mean; we refer to the error simply as RMSE. We observe that the random-points have lowest mean error when $\hat{\nu}/\nu$ is around 0.3, an underestimate. This mean grows mildly with $\hat{\nu}$, and the scatter across the trials is significant but steady. Toward lower $\hat{\nu}$, however, the mean RMSE grows quickly to twice

its minimum value and ultimately hits nearly three times that level. Coincident with this is an explosive growth of variability in the ensemble - a closer look at the distribution of cases shows there is a fraction of utterly failed reconstructions, with normalized errors greater than one. The RIPD points give a more consistent outcome in both mean and variability for the RMSE across the entire range of $\hat{\nu}/\nu$. Most notably, the mean RMSE is always lower than that of the random-points and thus the RIPD strategy on average improves the reconstruction error. The lower plot in Figure 2 confirms that for higher $\hat{\nu}/\nu$ a 15-20% reduction in RMSE is expected, whereas for low $\hat{\nu}/\nu$ this improvement is close to fifty percent. The variability is high however, and it must be kept in mind that only 60-70% of trials show an improvement. Overall, this result argues for setting $\hat{\nu}$ safely low - even zero when it is known that there is nonzero sensor noise - and using the RIPD strategy.

In two dimensions, for the DCT we assume no moves can be greater than 0.3 units Manhattan distance. The lower part of Figure 2 indicates a similar trend as for one dimension; as the sensor noise is underestimated, the RP paths deteriorate while the RIPD maintains a steady ensemble mean and standard deviation (not shown).

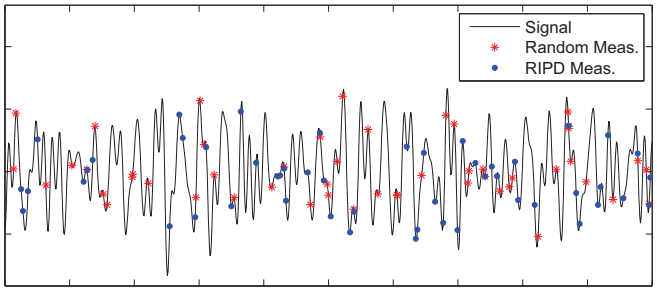

Fig. 1. A representative signal in the 1D case with discrete cosine transform. Fifty points are selected randomly, or designed according to the RIP, with one-hundred candidates considered for each addition.

B. Haar Wavelet Transform in Two Dimensions without Noise

Unlike the DCT, the Haar transform has a discrete domain, and almost all modes are zero on a fraction of it. This makes D very sparse and therefore amenable to speed improvements through proper programming. For instance in our example 32×32 image, $N = 1024$, but a given row of D has only 36 nonzero values, so that the number of nonzero scalar products in the row is 630, compared to the 523,776 products for a dense row. By maximizing re-use of the fixed part of the inner product set among a large group of candidates, we have been able to evaluate 200 candidates per second on a standard laptop computer. One might be tempted to apply this speed to a dynamic programming routine on the discrete state-space, but this attraction may be illusory because the inner product set for each given path-to-go still has to be maintained.

This example has no sensor noise, but we consider different step lengths. Figure 3 shows a test image derived from sea-surface temperature data, that has been explicitly sparsified.

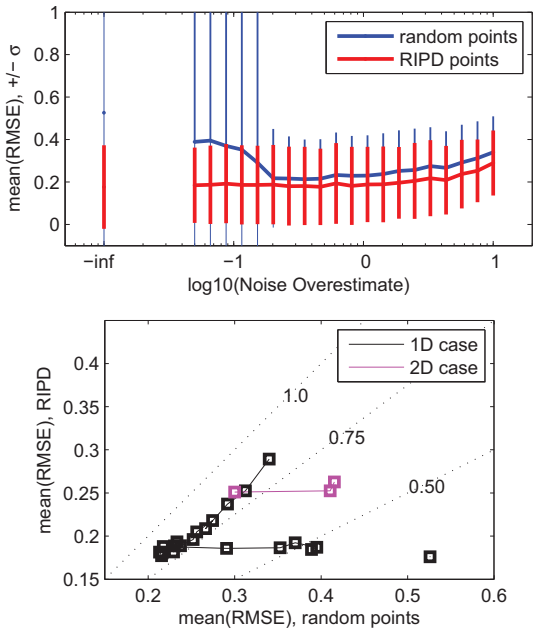

Fig. 2. DCT results summary. *Top:* 1D case, mean values of the random-point and RIPD ensembles, along with lines showing plus and minus one standard deviation for the ensemble, as a function of $\hat{\nu}/\nu$; the leftmost point represents $\hat{\nu} = 0$. *Bottom:* Alternate view of the mean values above for one dimension, and three similar points for two-dimensional case with $\hat{\nu}/\nu = [0, 0.05, 1]$, right to left.

The image that is to be reconstructed has fifty Haar elements and contains both large-scale and small-scale features, most of which come directly over from the full-resolution base image. Paths associated with the RP and the RIPD strategy illustrate that the RIPD is regularizing the points, in the classical sense of reducing discrepancy and dispersion, along the lines of quasi-Monte Carlo sequences such as the Halton and Hammersley.

For $m = 4S$, RMSE values are lowest when the kinematics are unconstrained; this is a natural analog to the standard CS recommendation of random points. Even in this case, however, Figure 4 indicates that design via RIPD will offer a modest but significant improvement over RP of about ten percent. When the stepsize constraints are turned on, the improvement offered by RIPD increases to about fifteen percent, with improvement occurring in seventy percent of trials. Looking at $m = 2S$, we expect and observe worse RMSE levels, with RIPD offering no significant improvement (points along the line of slope 1.0 on the graph). An exception occurs at stepsize five, however, where the random-point paths fail badly because sometimes the agent becomes localized in a corner and so does not visit or even cross the space. RIPD helps the agent to escape these conditions; improvements here occur in sixty percent of trials. At the other extreme, setting $m = 10S$ for the stepsize of ten reduces RMSE values, the reconstructions are very good, and RIPD again gives a mild improvement. We cannot achieve an exact reconstruction in general for the Haar wavelets because of limited support of the modes.

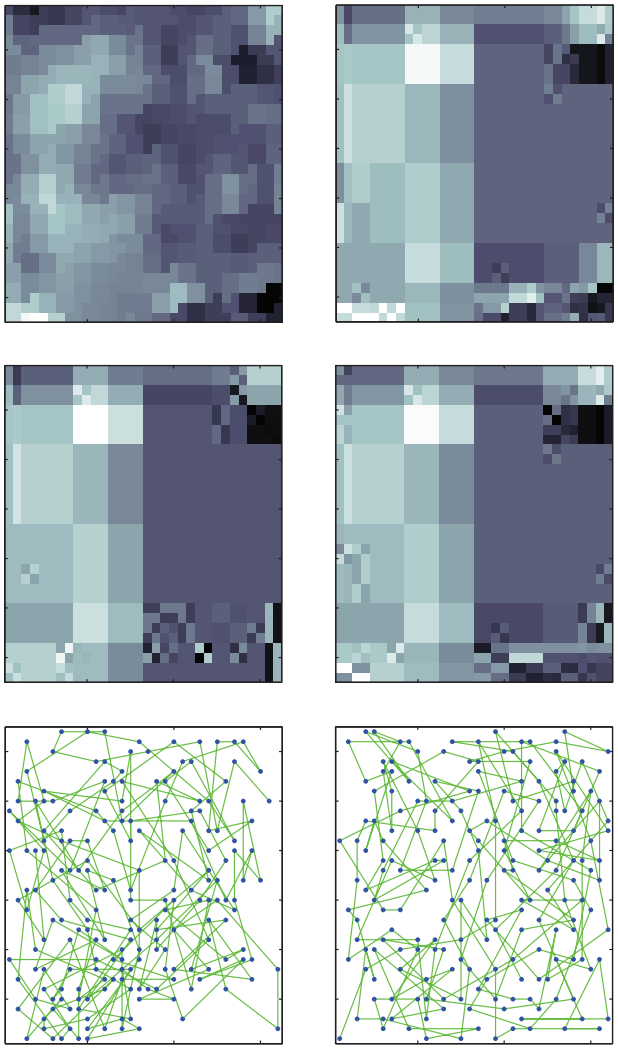

Fig. 3. *Top left:* Discretized image of a sea-surface temperature map. *Top right:* The sparsified SST map with S=50. *Middle left:* Typical reconstruction result using 200 points randomly selected, with maximum move 10 pixels. *Middle right:* Same for 200 RIPD points; the RMSE is 15% less than for random points. *Bottom left:* Random path associated with the reconstruction above. *Bottom right:* Same for RIPD path. *Note: the bounding box drawn is outside the feasible set of points.*

V. Conclusion

We have described a variation on standard compressive sensing, wherein we acknowledge the cost of transit and the desirability of interruptible CS missions. These and other aspects of CS data collection with mobile agents are interesting because they motivate new schemes for selecting points. Our main contribution in this paper was to apply a cheap rule for the incremental selection of points. A TSP path can efficiently visit a large set of such points, or they can be visited as generated; limited steplengths are then a simple way to reflect transit cost. Although the improvements in reconstruction error obtained through our RIPD points are modest, at 10-20% compared with random points, they appear to be robust; we encountered no conditions where the RIPD was outperformed.

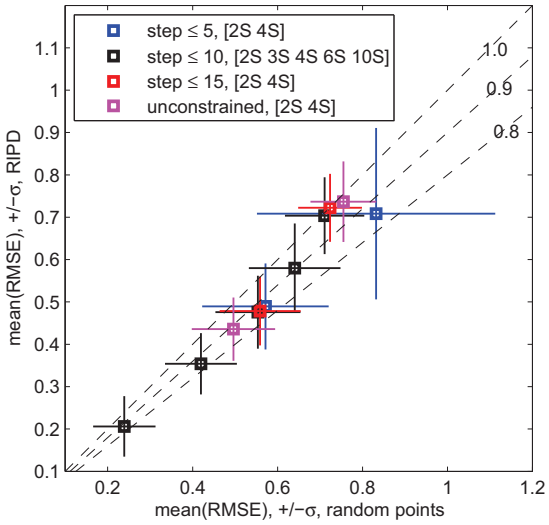

Fig. 4. Summary of 2D Haar results. Each color corresponds to a maximum step length (in pixels), with number of measurements m in square brackets. Mean values center on lines of ± 1 ensemble standard deviation. In all cases, increasing m reduces both random-point and RIPD reconstruction errors, as expected.

Acknowledgment

Work is supported by the Office of Naval Research, Grant N00014-09-1-0700.

References

[1] Candes, E.J., and T. Tau, Decoding by linear programming, IEEE Transactions on Information Theory, 51(12), 4203-4215, 2005.

[2] Candes, E.J., and M.B. Wakin, An Introduction to Compressive Sampling, IEEE Signal Processing Magazine, 21-30, March 2008.

[3] Choset, H., Coverage for Robotics - A Survey of Recent Results, Annals of Mathmatics and Artificial Intelligence, 31, 113-126, 2001.

[4] Camilli, R., C.M. Reddy, D.R. Yoerger, B.A.S. Van Mooy, M.V. Jakuba, J.C. Kinsey, C.P. McIntyre, S.P. Sylva, and J.V. Maloney, Tracking Hydrocarbon Plume Transport and Biodegradation at Deepwater Horizon, Science, 330(6001), 201-204, 2010.

[5] Smith, R.N., Y. Chao, P.P. Li, D.A. Caron, B.H. Jones, and G.S. Sukhatme, Planning and Implementing Trajectories for Autonomous Underwater Vehicles to Track Evolving Ocean Processes Based on Predictions from a Regional Ocean Model Int. Journal of Robotics Research, 29(12), 1475-1497, 2010.

[6] Elad, M., Optimized Projections for Compressed Sensing, IEEE Transactions on Signal Processing, 55(12), 5695-5702, 2007.

[7] Duarte-Carvajalino, J.M., and G. Sapiro, Learning to Sense Sparse Signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization, IEEE Trans. on Image Processing, 18(7), 1395-1408, 2009.

[8] Xu, J., Y. Pi, and Z. Cao, Optimized ProjectionMatrix for Compressive Sensing, EURASIP Journal on Advances in Signal Processing, 560349, 2010.

[9] Chen, C.-Y., and P.P. Vaidyanathan, Compressed Sensing in MIMO Radar, 42nd Asilomar Conference on Signals, Systems and Computers, 41 - 44, 2008.

[10] Strohmer, T., and B. Friedlander, Compressed Sensing for MIMO radar - Algorithms and Performance, 43rd Asilomar Conference on Signals, Systems and Computers, 464 - 468, 2009.

[11] Yu, Y., A. Petropulu and H.V. Poor, Measurement Matrix Design for Compressive Sensing Based MIMO Radar, IEEE Trans. on Signal Processing, to appear in 2011. Also available at http://arxiv.org/PS_cache/arxiv/pdf/1102/1102.5079v1.pdf.

[12] Karp, R.M., Probabilistic Analysis of Partitioning Algorithms for the Traveling-Salesman Problem in the Plane, Mathematics of Operations Research, 2(3), 209-224, 1977.

[13] Hummel, R., S. Poduri, F. Hover, U. Mitra, and G. Sukhatme, Mission Design for Compressive Sensing with Mobile Robots, Proc. Int. Conf. Robotics and Automation, Shanghai, China, 2011.
[14] Stein, D.M., An Asymptotic, Probabilistic Analysis of a Routing Problem,Mathematics of Operations Research, 3(2), 89-101, 1978.
[15] LaValle, S.M., and J.J. Kuffner, Randomized kinodynamic planning, Int. J. Robotics Research, 20(5), 378-400, 2001.
[16] Candes, E.J., Wakin, M.B., and Boyd, S.P., Enhancing Sparsity by Reweighted l(1) Minimization J. Fourier Analysis and Applications, 14(5-6), 877-905, 2008.