RECOVERY OF SPARSE PERTURBATIONS IN LEAST SQUARES PROBLEMS

Mert Pilanci[1] *and Orhan Arikan*[2]

[1]Department of Electrical Engineering and Computer Science, University of California, Berkeley, USA
[2]Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey
mert@eecs.berkeley.edu, oarikan@ee.bilkent.edu.tr

ABSTRACT

We show that the exact recovery of sparse perturbations on the coefficient matrix in overdetermined Least Squares problems is possible for a large class of perturbation structures. The well established theory of Compressed Sensing enables us to prove that if the perturbation structure is sufficiently *incoherent*, then exact or stable recovery can be achieved using linear programming. We derive sufficiency conditions for both exact and stable recovery using known results of ℓ_0/ℓ_1 equivalence. However the problem turns out to be more complicated than the usual setting used in various sparse reconstruction problems. We propose and solve an optimization criterion and its convex relaxation to recover the perturbation and the solution to the Least Squares problem simultaneously. Then we demonstrate with numerical examples that the proposed method is able to recover the perturbation and the unknown exactly with high probability. The performance of the proposed technique is compared in blind identification of sparse multipath channels.

***Index Terms*—** Compressed Sensing, Structured Total Least Squares, Structured Perturbations, Matrix Identification, Sparse Multipath Channels

I. INTRODUCTION

In various signal processing problems including deconvolution, channel identification and equalization, it is important to produce estimates for an unknown vector $\mathbf{x}_0$ from a set of measurements $\mathbf{y}_0$. Typically, a linear model is used to relate the unknowns to the available measurements: $\mathbf{y}_0 = \mathbf{A}_0\mathbf{x}_0$, where the matrix $\mathbf{A}_0 \in \mathbb{R}^{m\times n}$ describes a linear relationship. The well known Least Squares (LS) method for solving the overdetermined linear equations $\mathbf{A}_0\mathbf{x} = \mathbf{y}$ for $m > n$, is the Maximum Likelihood solution when the observations $\mathbf{y} = \mathbf{y}_0 + \mathbf{e}$ are subject to independent identically distributed Gaussian noise vector $\mathbf{e}$ and recovers $\mathbf{x}_0$ with some error [1]. Surprisingly, it was recently shown that, if $\mathbf{e}$ is sparse, exact recovery of $\mathbf{x}_0$ can be achieved for some classes of matrices $\mathbf{A}$ using linear programming [2]. However, in practice the elements of the coefficient matrix are also subject to errors since they may be results of some other measurements or obtained under some modeling assumptions. When there are errors in both, i.e., $\mathbf{A} = \mathbf{A}_0 + \mathbf{E}$ and $\mathbf{y} = \mathbf{y}_0 + \mathbf{e}$, the Total Least Squares (TLS) technique, which "corrects" the system with minimum perturbation so that it becomes consistent is widely used [3]. TLS also have Maximum Likelihood properties when the perturbations are zero mean i.i.d. Gaussian random variables.

It is known that the Total Least Squares problem is more ill-conditioned than the Least Squares problem because the amount of uncertainty greatly increases when we introduce perturbations in $\mathbf{A}$ [3]. Inspired by [2], we seek a TLS complement of that result and show in the next section that, if the perturbations $\mathbf{E}$ and $\mathbf{e}$ are sparse in some basis, then we may recover both the perturbations and the unknown $\mathbf{x}$ by knowing only the perturbed data $(\mathbf{A}_0 + \mathbf{E}, \mathbf{y}_0 + \mathbf{e})$.

II. NOVEL SPARSE PERTURBATION THEORY

Assume a true, consistent, overdetermined linear system of equations, $\mathbf{A}_0\mathbf{x}_0 = \mathbf{y}_0$, while the observed quantities are related via:

$$\mathbf{A} = \mathbf{A}_0 + \sum_{i=1}^{N} \mathbf{A}_i p_i \ , \qquad \mathbf{y} = \mathbf{y}_0 + \sum_{i=1}^{N} \mathbf{y}_i p_i \ , \tag{1}$$

where matrices $\mathbf{A}_i$ and vectors $\mathbf{y}_i$ are constants which form a possibly overcomplete basis for the perturbation $\mathbf{p} = [p_1 \dots p_N]^T$. Note that the above formulation allows the uncertainty over $\mathbf{A}$ and $\mathbf{y}$ to be correlated.

Case I: $\mathbf{x}_0$ is known Although the case where $\mathbf{x}_0$ is known might seem fictitious, there exists applications such as channel identification, which we design the signal $\mathbf{x}_0$ to sense the system matrix $\mathbf{A}$. This recovery scheme is known as Matrix Identification [4] and recently applied for Compressed Sensing Radar [5]. First we define the Restricted Isometry Constant (RIC) of a matrix. Then the following theorem demonstrates exact recovery of the perturbation using Basis Pursuit (BP).

Definition 1: For $s \in \mathbb{Z}^+$, define restricted isometry constant (RIC) δ_s of a matrix $\mathbf{\Phi}$ as the smallest nonnegative number such that

$$(1 - \delta_s)\,\|\mathbf{x}\|_2^2 \le \|\mathbf{\Phi}\mathbf{x}\|_2^2 \le (1 + \delta_s)\,\|\mathbf{x}\|_2^2 \tag{2}$$

holds for all vectors $\mathbf{x}$ which are s-sparse, i.e., have atmost s nonzero elements [6].

Theorem 2.1: (Exact Recovery) Let $\mathbf{p}$ be a *k*-sparse vector and δ_s be the RIC for,

$$\mathbf{\Phi}(\mathbf{x}_0) \triangleq \Big[\mathbf{A}_1\mathbf{x}_0 - \mathbf{y}_1 \Big| \ \cdots \ \Big| \mathbf{A}_N\mathbf{x}_0 - \mathbf{y}_N\Big]. \tag{3}$$

and $\delta_{2k} < \sqrt{2} - 1$. Then the following convex program:

$$\min \left\|\mathbf{p}'\right\|_1 \quad s.t. \ \ (\mathbf{A} - \sum_{i=1}^{N} \mathbf{A}_i p_i')\mathbf{x}_0 = \mathbf{y} - \sum_{i=1}^{N} \mathbf{y}_i p_i' \tag{4}$$

recovers $\mathbf{A}_0$ exactly.

Proof: Using (1) we get:

$$(\mathbf{A} - \sum_{i=1}^{N} \mathbf{A}_i p_i)\mathbf{x}_0 = \mathbf{y} - \sum_{i=1}^{N} \mathbf{y}_i p_i \ , \tag{5}$$

$$\mathbf{A}\mathbf{x}_0 - \mathbf{y} = \sum_{i=1}^{N} (\mathbf{A}_i\mathbf{x}_0 - \mathbf{y}_i) p_i \ , \tag{6}$$

$$\mathbf{A}\mathbf{x}_0 - \mathbf{y} = \mathbf{\Phi}(\mathbf{x}_0)\mathbf{p} \ , \tag{7}$$

When $\mathbf{x}_0$ is known, $\mathbf{\Phi}(\mathbf{x}_0) \in \mathbb{R}^{m\times N}$ is a known overcomplete dictionary satisfying Retricted Isometry Property (RIP) and the convex program (4) recovers the perturbation $\mathbf{p}$ as shown in [6] and therefore $\mathbf{A}_0$ is recovered exactly. If $N \leq m$ then recovery is trivial via directly solving (7) if $\mathbf{\Phi}(\mathbf{x}_0)$ is full rank. ■

Remark 1: It is straightforward to show that Toeplitz structured perturbations $\mathbf{A}_i$ with $\mathbf{y}_i = 0, \forall i$ result a Toeplitz $\mathbf{\Phi}(\mathbf{x}_0)$. It is known that deterministic Toeplitz matrices satisfy RIP of order $\mathcal{O}(n^\gamma)$ if $\mathbf{x}_0$ is deterministic and satisfies the PDACF property with γ [7]. If $\mathbf{x}_0$ is random, it is shown that Toeplitz matrices satisfy RIP of order k for many practical distributions, with probability exceeding $1-\exp(c_1\frac{n}{k^2})$ if $k \leq c_2\sqrt{\frac{n}{N}}$ where c_1, c_2 are constants [8].

Instead of RIP we can derive a sufficiency condition as follows:

Theorem 2.2: (Coherency of perturbations) Assume $\mathbf{A}_i\mathbf{x}_0 \neq \mathbf{y}_i \;\; \forall i$. If $\mu < \frac{1}{2k-1}$, where,

$$\mu \triangleq \max_{i\neq j} \frac{< \mathbf{A}_i\mathbf{x}_0 - \mathbf{y}_i, \mathbf{A}_j\mathbf{x}_0 - \mathbf{y}_j >}{\|\mathbf{A}_i\mathbf{x}_0 - \mathbf{y}_i\|_2 \|\mathbf{A}_j\mathbf{x}_0 - \mathbf{y}_j\|_2}, \tag{8}$$

then the convex program (4) recovers the perturbation exactly.

Corollary 2.3: If perturbations are unstructured as in the Total Least Squares problem then $\mathbf{A}_i$ are the standart basis and it is trivial to show that $\mu = 1$ and sparse perturbations can not be recovered exactly via any method. On the contrary, if perturbations are orthogonal, i.e., $\mathbf{A}_i^T\mathbf{A}_j = \mathbf{y}_i^T\mathbf{y}_j = \mathbf{A}_i^T\mathbf{y}_j = 0, \;\; \forall i \neq j$ then $\mu = 0$.

Case II: $\mathbf{x}_0$ is not known This is the general case examined in this paper and differs significantly from the usual setup of sparse recovery since the dictionary $\mathbf{\Phi}(\mathbf{x}_0)$ is unknown. A straightforward workaround is to employ the Least Squares solution $\mathbf{A}^\dagger\mathbf{y}$ of $\mathbf{x}_0$ and apply a regularized Basis Pursuit [9] with the estimate $\mathbf{\Phi}(\mathbf{A}^\dagger\mathbf{y})$. Using the recent results of [10] on dictionary perturbations we next prove that this scheme provides stable recovery under some conditions.

Theorem 2.4: (Stable Recovery) For a k-sparse $\mathbf{p}$, if RIC of $\mathbf{\Phi}(\mathbf{x}_0)$ satisfies:

$$\delta_{2k} < \frac{\sqrt{2}}{(1+2k\nu)^2} - 1 \;\;, \text{where} \;\; \nu \triangleq \frac{\max_i \;\; \left\|\mathbf{A}_i(\mathbf{A}^\dagger\mathbf{y} - \mathbf{x}_0)\right\|_2}{\min_j \;\; \|\mathbf{A}_j\mathbf{x}_0 - \mathbf{y}_j\|_2} \;\;, \tag{9}$$

and $k \leq m$, then the following convex program:

$$\min \left\|\mathbf{p}'\right\|_1 s.t. \left\|(\mathbf{A} - \sum_{i=1}^N \mathbf{A}_i p_i')\mathbf{A}^\dagger\mathbf{y} - (\mathbf{y} - \sum_{i=1}^N \mathbf{y}_i p_i')\right\|_2 \leq \boldsymbol{\epsilon} \tag{10}$$

provides stable recovery in the following sense:

$$\|\mathbf{p}^* - \mathbf{p}_0\|_2 \leq C\boldsymbol{\epsilon}, \tag{11}$$

where, $\mathbf{p}^*$ is the optimal solution of (10), C is a small constant and

$$\boldsymbol{\epsilon} \triangleq \left(k\nu\frac{\sqrt{1+\delta_k}}{\sqrt{1-\delta_k}} + \frac{\left\|\mathbf{A}(\mathbf{A}^\dagger\mathbf{y} - \mathbf{x}_0)\mathbf{y}\right\|_2}{\|\mathbf{r}\|_2}\right)\|\mathbf{r}\|_2 \;\;, \tag{12}$$

i.e., the error is in the order of the norm of $\mathbf{r} \triangleq \mathbf{A}\mathbf{x}_0 - \mathbf{y}$ which is the residual of the perturbed system.

Proof: Following the results of [10], we seek a bound for the worst case dictionary perturbation over k columns when we use the Least Squares estimate $\mathbf{x} = \mathbf{A}^\dagger\mathbf{y}$ in (7):

$$\begin{aligned}
&\frac{\max_{i_1,\ldots,i_k,i_p\neq i_q} \|[\mathbf{A}_{i_1}(\mathbf{x}-\mathbf{x}_0),\ldots,\mathbf{A}_{i_k}(\mathbf{x}-\mathbf{x}_0)]\|_2}{\max_{i_1,\ldots,i_k,i_p\neq i_q} \|[(\mathbf{A}_{i_1}\mathbf{x}_0-\mathbf{y}_{i_1}),\ldots,(\mathbf{A}_{i_k}\mathbf{x}_0-\mathbf{y}_{i_k})]\|_2} \qquad (13)\\
\leq \;& \frac{\max_{i_1,\ldots,i_k,i_p\neq i_q} \sqrt{k} \max_{i\in\{i_1,\ldots,i_k\}} \|\mathbf{A}_i(\mathbf{x}-\mathbf{x}_0)\|_2}{\max_{i_1,\ldots,i_k,i_p\neq i_q} \|[(\mathbf{A}_{i_1}\mathbf{x}_0-\mathbf{y}_{i_1}),\ldots,(\mathbf{A}_{i_k}\mathbf{x}_0-\mathbf{y}_{i_k})]\|_2}\\
\leq \;& \frac{\max_{i_1,\ldots,i_k,i_p\neq i_q} \sqrt{kR} \max_{i\in\{i_1,\ldots,i_k\}} \|\mathbf{A}_i(\mathbf{x}-\mathbf{x}_0)\|_2}{\max_{i_1,\ldots,i_k,i_p\neq i_q} \sum_{j\in\{i_1,\ldots,i_k\}} \|\mathbf{A}_j\mathbf{x}_0-\mathbf{y}_j\|_2}\\
\leq \;& k \; \frac{\max_i \;\; \|\mathbf{A}_i(\mathbf{x}-\mathbf{x}_0)\|_2}{\min_j \;\; \|\mathbf{A}_j\mathbf{x}_0-\mathbf{y}_j\|_2},
\end{aligned}$$

where $R = \mathit{Rank}[(\mathbf{A}_{i_1}\mathbf{x}_0 - \mathbf{y}_{i_1}),\ldots,(\mathbf{A}_{i_k}\mathbf{x}_0 - \mathbf{y}_{i_k})] \leq k$.

The perturbation in the left side of (7) is also bounded by, $\frac{\|\mathbf{A}(\mathbf{x}-\mathbf{x}_0)\|_2}{\|\mathbf{A}\mathbf{x}_0-\mathbf{y}\|}$. A straightforward application of Theorem 2 of [10] using the derived perturbation bounds completes the proof. ■

Remark 2: Note that the stability condition depends heavily on ν and consequently $\left\|\mathbf{A}^\dagger\mathbf{y} - \mathbf{x}_0\right\|$, which is known to scale with $\|\mathbf{A}\|_2^2 \left\|\mathbf{A}^\dagger\right\|_2^2$, the square of the condition number of $\mathbf{A}$ [11]. In particular, since δ_{2k} is nonnegative, the theorem requires $\nu < \frac{\sqrt[4]{2}-1}{2k}$ for stable recovery. Therefore, we conclude that two major limitations of perturbation recovery is the *ill-conditioning* of $\mathbf{A}$ and *coherency* of perturbations.

Remark 3: By using a corrective Min-Min approach that will be introduced next, the performance of this estimator may be improved significantly.

III. PROPOSED ESTIMATOR WHEN $\mathbf{x}_0$ IS NOT KNOWN

The following double minimization is proposed for joint estimation of the sparse perturbation $\mathbf{p}$ and unknown $\mathbf{x}_0$:

$$P_0 : \min_{\mathbf{x}} \min_{\|\mathbf{p}\|_0 = k} \left\|(\mathbf{A} - \sum_{i=1}^N \mathbf{A}_i p_i)\mathbf{x} - (\mathbf{y} - \sum_{i=1}^N \mathbf{y}_i p_i)\right\|_2$$

III-A. Alternating Minimizations Algorithm to solve P_0

When $\mathbf{p}$ is fixed the problem reduces to a simple Least Squares problem which can be solved via the pseudoinverse. If $\mathbf{x}$ is fixed then there exists many algorithms to solve for a sparse $\mathbf{p}$ [12]. Therefore a local optimum can be found using an alternating minimizations algorithm [13] where we chose Orthogonal Matching Pursuit (OMP) [14] in the intermediate step for its simplicity:

Algorithm 1. Alternating Minimizations for P_0

$\mathbf{x}^0 \leftarrow \mathbf{A}^\dagger\mathbf{y}$, $\mathbf{p}^0 \leftarrow \mathbf{0}$, $k \leftarrow 0$
while $\left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\| > \Delta$ **do**
  $\mathbf{p}^{k+1} \leftarrow \arg\min_{\|\mathbf{p}\|_0=k} \left\|\mathbf{A}\mathbf{x}^k - \mathbf{y} - \mathbf{\Phi}(\mathbf{x}^k)\mathbf{p}^k\right\|$
  (using OMP)
  $\mathbf{x}^{k+1} \leftarrow (\mathbf{A} - \sum_{i=1}^N \mathbf{A}_i p_i)^\dagger(\mathbf{y} - \sum_{i=1}^N \mathbf{y}_i p_i)$
  $k \leftarrow k+1$
end while
$\hat{\mathbf{A}} \leftarrow (\mathbf{A} - \sum_{i=1}^N \mathbf{A}_i p_i^k)$, $\hat{\mathbf{y}} \leftarrow (\mathbf{A} - \sum_{i=1}^N \mathbf{y}_i p_i^k)$, $\hat{\mathbf{x}} \leftarrow \mathbf{x}^k$

III-B. Convex Relaxation of the Proposed Estimator

If the constraint on $\mathbf{p}$ is relaxed to l_1 norm as follows, faster gradient based techniques can be used to solve the problem since the objective of P_0 is convex in both $\mathbf{x}$ and $\mathbf{p}$ (but not jointly):

$$P_1 : \min_{\mathbf{x}} \min_{\|\mathbf{p}\|_1 \leq t} \left\| (\mathbf{A} - \sum_{i=1}^{N} \mathbf{A}_i p_i)\mathbf{x} - (\mathbf{y} - \sum_{i=1}^{N} \mathbf{y}_i p_i) \right\|_2 .$$

First define the following matrix functions:

Definition 2: Let,

$$\mathbf{A}(\mathbf{p}) \triangleq \mathbf{A} - \sum_{i=1}^{N} \mathbf{A}_i p_i \quad , \qquad \mathbf{y}(\mathbf{p}) \triangleq \mathbf{y} - \sum_{i=1}^{N} \mathbf{y}_i p_i \quad . \tag{14}$$

Assuming $\mathbf{A}(\mathbf{p})$ is of full column rank for $\|\mathbf{p}\|_1 \leq t$, the outer minimization of P_1 can be carried out analytically as:

$$\min_{\|\mathbf{p}\|_1 \leq t} \min_{\mathbf{x}} \|\mathbf{A}(\mathbf{p})\mathbf{x} - \mathbf{y}(\mathbf{p})\|_2 = \min_{\|\mathbf{p}\|_1 \leq t} \left\| \mathbf{P}_{\mathbf{p}}^{\perp} \mathbf{y}(\mathbf{p}) \right\|_2 \tag{15}$$

where $\mathbf{P}_{\mathbf{p}}^{\perp} \triangleq \mathbf{I} - \mathbf{A}(\mathbf{p})\mathbf{A}(\mathbf{p})^{\dagger}$ is the projector matrix of the subspace perpendicular to the Range($\mathbf{A}(\mathbf{p})$). Let $\mathbf{y}(\mathbf{p})^{\perp} \triangleq \mathbf{P}_{\mathbf{p}}^{\perp}\mathbf{y}(\mathbf{p})$ and $\mathbf{x}_{\mathbf{p}} \triangleq \mathbf{A}(\mathbf{p})^{\dagger}\mathbf{y}(\mathbf{p})$, the authors prove the following in [15]:

$$\frac{1}{2}\frac{\partial}{\partial p_i} \left\| \mathbf{P}_{\mathbf{p}}^{\perp} \mathbf{y}(\mathbf{p}) \right\|_2^2 = \left\langle \mathbf{y}(\mathbf{p})^{\perp},\, \mathbf{A_i x_p} - \mathbf{y_i} \right\rangle, \tag{16}$$

which makes P_1 solvable using fast gradient based techniques such as the following:

Coordinate Gradient Descent (CGD): CGD is a gradient based algorithm to solve l_1 constrained optimization problems [16]. The following adaptation of CGD provides a solution to P_1:

Algorithm 2 CGD for P_1

$\mathbf{p}^0 \leftarrow \mathbf{0}$, $k \leftarrow 0$
while $\left\|\mathbf{p}^k - \mathbf{p}^{k-1}\right\| > \Delta$ **do**
$\quad l \leftarrow \arg\min_i \left| \langle \mathbf{y}(\mathbf{p})^{\perp},\, \mathbf{A_i x_p} - \mathbf{y_i} \rangle \right|$
$\quad \mathbf{c} \leftarrow \mathbf{0}$, $\quad \mathbf{c}_k \leftarrow -sign(< \mathbf{y}(\mathbf{p})^{\perp},\, \mathbf{A_l x_p} - \mathbf{y_l}) >$
$\quad \hat{\lambda} \leftarrow \arg \min_{\lambda \in [0,1]} \left\| \mathbf{P}^{\perp}_{\mathbf{p}^k + \lambda(\mathbf{c} - \mathbf{p}^k)} \mathbf{y}(\mathbf{p}^k + \lambda(\mathbf{c} - \mathbf{p}^k)) \right\|_2$
$\quad \mathbf{p}^{k+1} \leftarrow \mathbf{p} + k + \hat{\lambda}(\mathbf{c} - \mathbf{p} + k)$
$\quad k \leftarrow k + 1$
end while
$\hat{\mathbf{A}} \leftarrow (\mathbf{A} - \sum_{i=1}^{N} \mathbf{A}_i p_i^k)$, $\ \hat{\mathbf{y}} \leftarrow (\mathbf{A} - \sum_{i=1}^{N} \mathbf{y}_i p_i^k)$, $\ \hat{\mathbf{x}} \leftarrow \mathbf{x}^k$

Remark 4: The exact optimization over λ is non-convex. However it can be accurately approximated via the following:

$$\begin{aligned} & \arg \min_{\lambda \in [0,1]} \left\| \mathbf{P}^{\perp}_{\mathbf{p} + \lambda(\mathbf{c} - \mathbf{p})} \mathbf{y}(\mathbf{p} + \lambda(\mathbf{c} - \mathbf{p})) \right\|_2 \\ \approx \quad & \arg \min_{\lambda \in [0,1]} \|\mathbf{A}(\mathbf{p} + \lambda(\mathbf{c} - \mathbf{p}))\mathbf{x} - \mathbf{y}(\mathbf{p} + \lambda(\mathbf{c} - \mathbf{p}))\|_2 \\ = \quad & \begin{cases} 0 & \alpha(\mathbf{A}^{\dagger}\mathbf{y}) \leq 0 \\ \alpha(\mathbf{A}^{\dagger}\mathbf{y}) & 0 < \alpha(\mathbf{A}^{\dagger}\mathbf{y}) < 1 \\ 1 & \alpha(\mathbf{A}^{\dagger}\mathbf{y}) > 1 \end{cases} \end{aligned}$$

$$\text{where } \alpha(\mathbf{x}) \triangleq \frac{(\mathbf{A}(\mathbf{p})\mathbf{x} - \mathbf{y}(\mathbf{p}))^T \sum_i (\mathbf{A}_i \mathbf{x} - \mathbf{y}_i)(\mathbf{c} - \mathbf{p})}{\left\| \sum_i (\mathbf{A}_i \mathbf{x} - \mathbf{y}_i)(\mathbf{c} - \mathbf{p}) \right\|_2^2} \quad .$$

IV. NUMERICAL RESULTS AND APPLICATIONS

IV-A. Probability of Exact Recovery

For the case $\mathbf{x}_0$ is unknown, the exact recovery of perturbation may seem hopeless. However we demonstrate that exact recovery can be achieved with a high probability with the proposed estimator if the *overdetermination ratio* $\frac{m}{n}$ of the matrix is sufficiently large. A Toeplitz matrix with random elements $\mathbf{A}_0$ is perturbed by preserving the structure with k-sparse perturbations $\mathbf{p}$ and P_0 is solved to recover the perturbation. The empirical probability of exact recovery in 100 trials versus the ratio m/n is shown in Figure 1(a). And the probability of exact recovery is examined in the $(\frac{m}{n}, k)$ plane in Figure 1(b).

IV-B. Blind Identification of Sparse Multipath Channels

Consider a channel model which consists of N_p paths with attenuations a_i, delays n_i and doppler shifts ν_i :

$$y[n] \quad = \quad \sum_{i=1}^{N_p} a_i x_0[n - n_i] e^{j2\pi \frac{\nu_i}{d}} + w[n] \quad , \tag{17}$$

which can be written more compactly as, $\mathbf{y} = \mathbf{H}\mathbf{x_0} + \mathbf{w}$, where $\mathbf{w}$ is circularly symmetric Gaussian white noise. Here we consider the joint estimation of the channel and its input following a training session that provided a channel estimate $\mathbf{H}$ with N_p paths. Since the paths are usually sparse in delay-doppler domain [17], the problem turns out to be a sparse perturbation recovery problem over a discretized delay-doppler domain with bins of length $\Delta\nu = \frac{1}{n}$, $\ \Delta\tau = \frac{1}{B}$ where B is the bandwidth of $x_0[n]$. To simplify the development, we define $N = md$ structure matrices as the following basis of time-frequency shifts [5]:

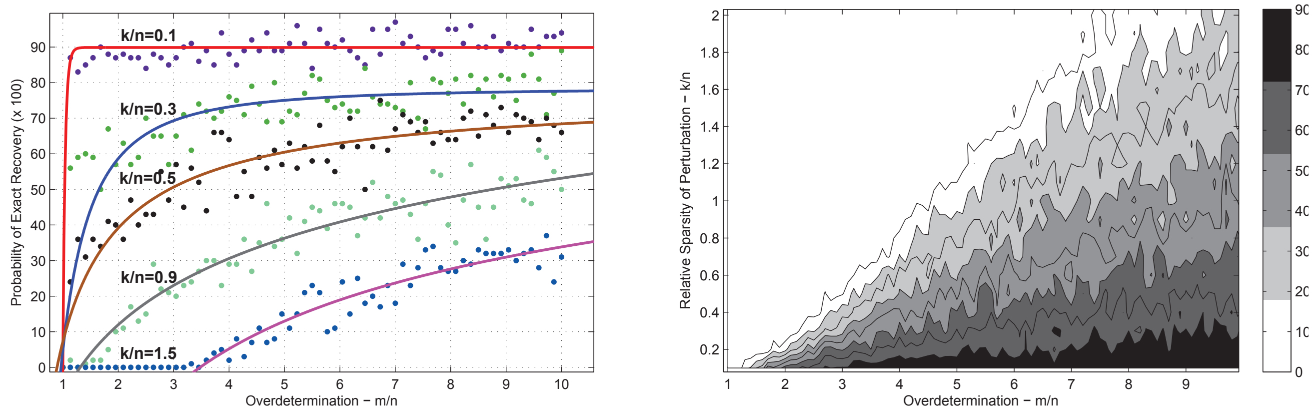
$$\mathbf{H}_{kl} = diag([1 \ e^{j\frac{2\pi}{d}k} \ \ldots \ e^{j\frac{2\pi}{d}km}])\mathbf{R}_l \quad , \tag{18}$$

where $\mathbf{R}_l$ is a matrix whose l^{th} subdiagonal entries are 1 and the rest is zero, effectively performing *shift by* l operation.

Note that $\mathbf{H}_{kl}$'s have Toeplitz structure and generate sufficiently incoherent perturbations depending on $x_0[n]$ as we outline in Remark 1. A simulation study is done to demonstrate the performance of proposed solver P_1 where $\mathbf{x}_0$ is selected as a random ± 1 sequence and assumed unknown. 1, 3 and 5 more paths with unknown attenuations, delays and doppler shifts are added respectively to a known channel $\mathbf{H}$. The parameter t is selected such that the perturbation sparsity matches the number of unknown channels. In Figures 2 and 3, the Basis Pursuit approach that we described in (4) where $\mathbf{x}_0$ is known is compared in terms of doppler and delay estimation error defined by $\sqrt{\frac{\nu}{d\Delta\nu}}$ and $\sqrt{\frac{\tau}{d\Delta\tau}}$ by averaging 100 realizations of noise in 30 SNR levels. Although $\mathbf{x}_0$ is unknown, the proposed scheme outperforms BP in terms of perturbation recovery and successfully estimates both the input sequence by identifying unknown paths in the channel.

V. CONCLUSION

We showed that the exact or stable recovery of sparse perturbations in Least Squares problems is achievable under some conditions. It is found that, ill-conditioning of the matrix and coherence of perturbations are the limitations of perturbation recovery. If $\mathbf{x}_0$ is known, exact recovery can be achieved. For the case when $\mathbf{x}_0$ is unknown, we proposed an optimization scheme and its convex relaxation to recover perturbations. The numerical examples show that the empirical probability of exact recovery is high for reasonable overdetermination ratios and it has superior performance when applied to identification of sparse multipath channels.



(a) Empirical probability of exact recovery versus overdetermination ratio of the proposed estimator P_0 for Toeplitz structured sparse perturbations.

(b) Empirical probability of exact recovery in the $(\frac{m}{n}, \frac{k}{n})$ plane.

Fig. 1. Empirical probability of exact recovery for the case where $\mathbf{x}_0$ is unknown.

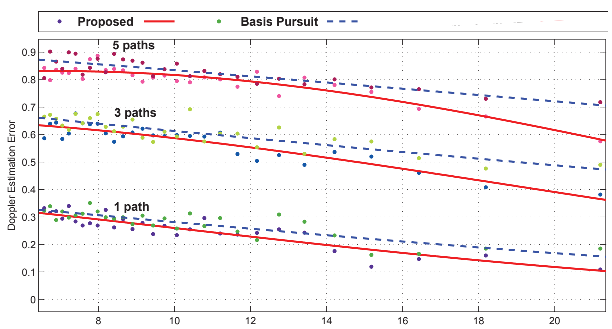


Fig. 2. Normalized Doppler Estimation Error.

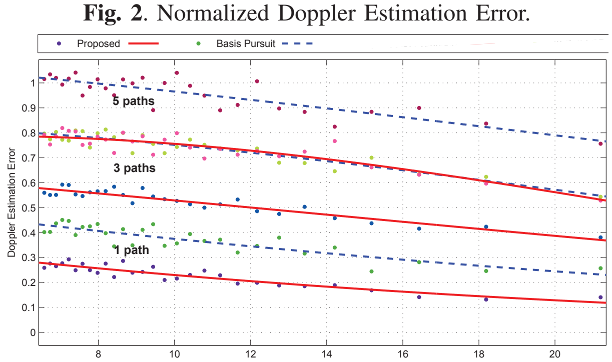


Fig. 3. Normalized Delay Estimation Error.

VI. REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.*, Prentice-Hall, 1993.

[2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[3] G.H. Golub and F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 883–893, Dec. 1980.

[4] G.E. Pfander, H. Rauhut, and J. Tanner, "Identification of matrices having a sparse representation," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5376–5388, Nov 2008.

[5] M.A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun 2009.

[6] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *C. R. Math.*, vol. 346, no. 9–10, pp. 589–592, May 2008.

[7] V. Saligrama, "Deterministic designs with deterministic guarantees: Toeplitz compressed sensing matrices, sequence designs and system identification," *arXiv:0806.4958*, Jun 2008.

[8] W. Bajwa, J.D. Haupt, G.M. Raz, S.J. Wright, and R.D. Nowak, "Toeplitz-structured compressed sensing matrices," in *IEEE/SP 14th Workshop on Statistical Signal Processing, SSP 07, pp. 294–298, Aug, 2007.*

[9] E. J. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 120–1223, Aug 2006.

[10] M. Herman and T. Strohmer, "General deviants: an analysis of perturbations in compressed sensing," *arXiv:0907.2955*, Feb 2009.

[11] G. W. Stewart, "On the perturbation of pseudo-inverses, projections and linear least squares problems," *SIAM Review*, vol. 19, no. 4, pp. 634–662, Oct 1977.

[12] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *Journal of Fourier Analysis and Applications*, vol. 51, no. 1, pp. 34–81, Feb 2009.

[13] M. Pilanci, O. Arikan, B. Oguz, and M.C. Pinar, "Structured least squares with bounded data uncertainties," in *IEEE Int. Conf. Acoust. Speech Sign. Processing (ICASSP)*, 2009.

[14] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 1977.

[15] M. Pilanci, O. Arikan, and M.C. Pinar, "Structured least squares problems and robust estimators," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2453–2465, May 2010.

[16] K. Yongdai and K. Jinseog, "Gradient lasso for feature selection," in *Proceedings of the twenty-first international conference on Machine learning, July, 2004.*

[17] W. Bajwa, A. Sayeed, and R. Nowak, "Sparse multipath channels: Modeling and estimation," in *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, pp. 320-325, 2008.*