ON THE THEORETICAL ANALYSIS OF CROSS VALIDATION IN COMPRESSIVE SENSING

Jinye Zhang [†]   *Laming Chen* [†]   *Petros T. Boufounos* [⋆]   *Yuantao Gu* [†]

[†] State Key Laboratory on Microwave and Digital Communications
Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA
[⋆] Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, USA

ABSTRACT

Compressive sensing (CS) is a data acquisition technique that measures sparse or compressible signals at a sampling rate lower than their Nyquist rate. Results show that sparse signals can be reconstructed using greedy algorithms, often requiring prior knowledge such as the signal sparsity or the noise level. As a substitute to prior knowledge, cross validation (CV), a statistical method that examines whether a model overfits its data, has been proposed to determine the stopping condition of greedy algorithms. This paper analyses cross validation in a general compressive sensing framework. Furthermore, we provide both theoretical analysis and numerical simulations for a cross-validation modification of orthogonal matching pursuit, referred to as OMP-CV, which has good performance in sparse recovery.

Index Terms— Compressed sensing, signal reconstruction, cross validation, orthogonal matching pursuit

1. INTRODUCTION

Compressive sensing (CS) is a data acquisition technique that measures sparse or compressible signals at sampling rate close to their intrinsic information rate rather than the Nyquist rate [2, 3]. Broadly speaking, it consists of two main building blocks: encoding an N-dimensional k-sparse signal $\mathbf{x}$ by computing its $M \ll N$ linear projections and decoding the signal using various sparse recovery methods, such as linear programming [4–6] or greedy algorithms [7–11].

While greedy algorithms are often preferred for their low computational complexity, most of them require prior information—such as signal sparsity or noise level—for accurate reconstruction, without which the reconstructed signal may fail to completely reconstruct the signal or overfit the noise [17]. Such information not available in practice, cross validation (CV) [12–16], a statistical method that examines whether a model overfits the data, has been proposed as an alternative solution [17], where, in general, measurements are separated into reconstruction measurements and CV measurements; the former are used to reconstruct the signal via a greedy algorithm, while the latter to compute the stopping criterion. It has also been shown that CV can be used to estimate the recovery error. Specifically, using the Johnson-Lindenstrauss lemma, we can derive theo-

retical bounds on the estimation error, quantifying how the number of CV measurements m_{cv} influences the estimation quality [18].

In this paper we provide an analysis of CV performance in general CS problems. Furthermore, we analyse a variation of the orthogonal matching pursuit (OMP) algorithm, called OMP-CV, which exploits CV to determine the recovered signal sparsity.

To analyse CV in general CS problems, we need to address the following two problems. First, given a recovered signal, we are interested in determining how close it is to the original input signal. Second, given two candidate recovered signals, we are interested in determining which of the two has lower recovery error.

Both questions could be answered using CV. Given a signal, the recovery error ε_{x} can be estimated by its CV residual ϵ_{cv}; we attempt to determine the accuracy and probability the CV residual could provide bounds on the recovery error. Furthermore, given two signals, the comparison of two recovery errors can be estimated by their respective CV residuals; we are interested in the probability the comparison of the error and the CV residual are consistent with each other. In other words, these problems could be formulated as:

1. (Recovery error estimation) Given a recovered signal, with what accuracy and what probability could its CV residual provide bounds on its recovery error?

2. (Recovery error comparison) Given two recovered signals, with what probability could the comparison of their CV residuals correctly reflect their recovery errors?

These problems are referred to as *general CV problems* in the remainder of the article. To analyse both problems, we first calculate the probability distribution of CV residuals. By transforming this distribution into inequalities that hold with certain probability, we directly answer the questions above. Our results for both problems are given in Section 3.

We also analyse the OMP-CV algorithm, described in Table 1. In this analysis we refer to the algorithm output, which is the recovered signal with the smallest CV residual, as the *OMP-CV output*. The recovered signal with the smallest recovery error is referred to as *oracle optimum*.

Our main result shows that the recovery error of the OMP-CV output is very close to that of the oracle optimum with high probability given that the oracle optimum recovers all indices in the support set of the signal. Note that OMP-CV includes two almost separate procedures: one is reconstructing the signal using OMP, the other is estimating the recovery error by CV. To achieve the main result, we first analyse the internal structure of two recovered signals acquired in different iterations of OMP. We then study how their CV residuals correspond to their recovery errors using the techniques we developed for the general CV problem. Finally we generalize the recovery error comparison between two recovered signals, as generated

by the OMP iterations, to the comparison of all recovered signals. Thus we can estimate how close is the OMP-CV output to the oracle optimum.

The reminder of the paper is organized as follows. Section 2 formulates the problem and describes the OMP-CV algorithm. Section 3 gives the analysis of CV, while Section 4 gives the analysis of OMP-CV. Numerical simulations are given in Section 5. Finally, Section 6 discusses our findings and concludes.

2. BACKGROUND

2.1. Notation and Problem Formulation

We consider an unknown k-sparse signal $\mathbf{x} \in \mathbb{R}^N$ observed using M linear measurements corrupted by additive noise. Let T be its support set, using $|T| = k$ to denote the size of T. The vector $\mathbf{x}_T$ contains the coefficients of $\mathbf{x}$ indexed by T. To implement the CV-based modification, we separate the M by N sampling matrix to a reconstruction matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ and a CV matrix $\mathbf{A}_{\mathrm{cv}} \in \mathbb{R}^{m_{\mathrm{cv}} \times N}$. Measurements are also separated accordingly, to reconstruction measurements $\mathbf{y} \in \mathbb{R}^m$ and CV measurements $\mathbf{y}_{\mathrm{cv}} \in \mathbb{R}^{m_{\mathrm{cv}}}$.

Our analysis only considers Gaussian sampling matrices and additive Gaussian noise, i.e.,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \qquad \mathbf{n} = \sigma_{\mathrm{n}} \mathbf{a}_{\mathrm{n}},$$
$$\mathbf{y}_{\mathrm{cv}} = \mathbf{A}_{\mathrm{cv}}\mathbf{x} + \mathbf{n}_{\mathrm{cv}}, \qquad \mathbf{n}_{\mathrm{cv}} = \sigma_{\mathrm{n}} \mathbf{a}_{\mathrm{cv,n}},$$

where elements of $\mathbf{A}$, $\mathbf{A}_{\mathrm{cv}}$, $\mathbf{a}_{\mathrm{n}}$, and $\mathbf{a}_{\mathrm{cv,n}}$ are i.i.d normally distributed with mean zero and variance $1/m$ so that the sensing matrix will have unit column norm. For notational simplicity, we also define:

Definition 1. *(Generalized sampling matrix and input signal): Let $\mathbf{A}_{\mathrm{g}} \triangleq [\mathbf{A}, \mathbf{a}_{\mathrm{n}}]$ and $\mathbf{x}_{\mathrm{g}} \triangleq [\mathbf{x}', \sigma_{\mathrm{n}}]'$, then*

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} = \mathbf{A}_{\mathrm{g}}\mathbf{x}_{\mathrm{g}}, \tag{1}$$

where $\mathbf{A}_{\mathrm{g}}$ is called the generalized sampling matrix, and $\mathbf{x}_{\mathrm{g}}$ is called the generalized input signal.

In describing the OMP algorithm, we use $\hat{\mathbf{x}}^p$ to denote the recovered signal in the p-th iteration, and T^p to denote its support set. The difference between the recovered signal $\hat{\mathbf{x}}^p$ and the input signal $\mathbf{x}$ is denoted using $\Delta\mathbf{x}^p$ and the corresponding recovery error using $\varepsilon_{\mathrm{x}}^p$. The generalized versions of $\Delta\mathbf{x}^p$ and $\varepsilon_{\mathrm{x}}^p$ are $\Delta\mathbf{x}_{\mathrm{g}}^p$ and $\varepsilon_{\mathrm{g}}^p$, respectively, and ϵ_{cv}^p denotes the CV residual of $\hat{\mathbf{x}}^p$. In other words,

$$\Delta\mathbf{x}^p \triangleq \mathbf{x} - \hat{\mathbf{x}}^p, \qquad \varepsilon_{\mathrm{x}}^p \triangleq \|\Delta\mathbf{x}^p\|_2^2,$$
$$\Delta\mathbf{x}_{\mathrm{g}}^p \triangleq [(\Delta\mathbf{x}^p)', \sigma_{\mathrm{n}}]', \qquad \varepsilon_{\mathrm{g}}^p \triangleq \|\Delta\mathbf{x}_{\mathrm{g}}^p\|_2^2,$$
$$\epsilon_{\mathrm{cv}}^p \triangleq \|\mathbf{y}_{\mathrm{cv}} - \mathbf{A}_{\mathrm{cv}}\hat{\mathbf{x}}^p\|_2^2.$$

To make the analysis more clear, we emphasize that in this paper, the input signal is considered as deterministic while the sampling matrix and noise are random. Without loss of clarity and for notational simplicity, the random variables and their realizations are denoted by same notation.

2.2. OMP-CV: Algorithm Description

The OMP-CV algorithm, described in Table 1, is a noise- and sparsity-robust greedy recovery algorithm that combines OMP and CV [17]. The path it follows can be interpreted intuitively: as the signal estimate improves, the recovery error and CV residual decrease; when the recovered signal starts to overfit the noise, the

Table 1. OMP-CV Algorithm

Input: Reconstruction matrix $\mathbf{A}$, CV matrix $\mathbf{A}_{\mathrm{cv}}$, reconstruction measurements $\mathbf{y}$, CV measurements $\mathbf{y}_{\mathrm{cv}}$, iteration times d;
Output: The recovered signal $\hat{\mathbf{x}}$.
Initialization: Set $p = 1$, $\epsilon_{\mathrm{cv}}^0 = \|\mathbf{y}_{\mathrm{cv}}\|_2^2$;
Repeat: Compute $\hat{\mathbf{x}}^p$ using an OMP iteration; Compute $\epsilon_{\mathrm{cv}}^p = \|\mathbf{A}_{\mathrm{cv}}\hat{\mathbf{x}}^p - \mathbf{y}_{\mathrm{cv}}\|_2^2$; Increment p by 1;
Until: $p \geq d$ Compute $o_{\mathrm{cv}} = \underset{p}{\operatorname{argmin}}\, \epsilon_{\mathrm{cv}}^p$;
Return: $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{o_{\mathrm{cv}}}$.

recovery error increases. The CV residual detects the change by starting to increase simultaneously. That is, CV residuals provide a good estimate for recovery errors and, further, the OMP-CV output is close to the oracle optimum.

There are several advantages of OMP-CV. First, it does not require prior information such as noise level or sparsity. Instead, only the maximum number of iterations, d, is required as input[1]. This can be set two or three times larger than the sparsity k without hurting recovery performance. Second, the algorithm provides an estimate of the recovery error in its residual. Additionally, by properly setting m_{cv}, the recovery performance of OMP-CV competes with that of OMP when combined with accurate information on the noise level.

Earlier work has provided experimental support on the recovery performance of OMP-CV [17, 18]. However, to the best of our knowledge, theoretical analysis on this algorithm does not exist. Our work presents such analysis in Section 4. In addition, we also provide numerical simulations in Section 5 to support our theoretical results.

3. CROSS VALIDATION IN COMPRESSIVE SENSING

This section describes our results for general CV problems, as described in Section 1. We start with calculating the probability distribution of ϵ_{cv}, i.e.

Lemma 1. *Let $\hat{\mathbf{x}}$ be a recovered signal and ε_{x} be its recovery error. Provided that m_{cv} is sufficiently large [2], then*

$$\epsilon_{\mathrm{cv}} = \|\mathbf{y}_{\mathrm{cv}} - \mathbf{A}_{\mathrm{cv}}\hat{\mathbf{x}}\|_2^2 \sim \mathcal{N}(\mu, \sigma^2), \tag{2}$$

where $\mu = \frac{m_{\mathrm{cv}}}{m}(\varepsilon_{\mathrm{x}} + \sigma_{\mathrm{n}}^2)$, and $\sigma^2 = \frac{2m_{\mathrm{cv}}}{m^2}(\varepsilon_{\mathrm{x}} + \sigma_{\mathrm{n}}^2)^2$.

The condition "provided that m_{cv} is sufficiently large" is used in one approximation step of the proof which uses the Central Limit Theorem (CLT). The actual probability distribution of ϵ_{cv} converges absolutely to (2) as m_{cv} increases and the approximation error becomes negligible very fast. In practical cases the number of CV measurements m_{cv} is often tens or hundreds and the approximation is very good. The precision of the approximation is also supported by the simulation result in Section 5.1. The same condition is also

---
[1] Note that d cannot be greater than m, since regular OMP will produce zero residual after that and conclude.

[2] An m_{cv} exceeding tens, which is common in CS, meets the requirement.

required in Lemma 2, Theorem 1, and Theorem 2 for the same reason.

One immediate consequence of Lemma 1 is that ϵ_{cv} can be used to provide an estimate of ε_x in the form of an inequality that holds with certain probability. In particular:

Theorem 1. *(Recovery error estimation): Provided that m_{cv} is sufficiently large, with probability* $\mathrm{erf}(\frac{\lambda}{\sqrt{2}})$ *the following holds*

$$h(\lambda, +)\epsilon_{cv} - \sigma_n^2 \leq \varepsilon_x \leq h(\lambda, -)\epsilon_{cv} - \sigma_n^2, \quad (3)$$

where $h(\lambda, \pm)$ is a function related to λ defined as

$$h(\lambda, \pm) \triangleq \frac{m}{m_{cv}} \frac{1}{1 \pm \lambda\sqrt{\frac{2}{m_{cv}}}}, \quad (4)$$

and $\mathrm{erf}(u)$ is the error function of normal distribution,

$$\mathrm{erf}(u) \triangleq \frac{1}{\sqrt{\pi}} \int_{-u}^{u} e^{-t^2} dt. \quad (5)$$

Theorem 1 provides an answer to Prob. 1 bounding the recovery error ε_x by the interval

$$\left[h(\lambda, +)\epsilon_{cv} - \sigma_n^2, h(\lambda, -)\epsilon_{cv} - \sigma_n^2 \right]. \quad (6)$$

The difference of the upper bound and the lower bound,

$$\frac{m}{m_{cv}} \frac{2\lambda\sqrt{2}}{\sqrt{m_{cv}} - \frac{2\lambda^2}{\sqrt{m_{cv}}}} \epsilon_{cv}, \quad (7)$$

is roughly proportional to $1/m_{cv}^{2/3}$ and becomes tighter as m_{cv} increases.

For Prob. 2, we compute the distribution of $\Delta\epsilon_{cv} = \epsilon_{cv}^p - \epsilon_{cv}^q$.

Lemma 2. *Let $\hat{\mathbf{x}}^p$ and $\hat{\mathbf{x}}^q$ be two recovered signals, ε_g^p and ε_g^q be the their generalized recovery error, and*

$$\rho_g \triangleq \frac{\langle \Delta\mathbf{x}_g^p, \Delta\mathbf{x}_g^q \rangle}{\|\Delta\mathbf{x}_g^p\|_2 \|\Delta\mathbf{x}_g^q\|_2}. \quad (8)$$

Provided that m_{cv} is sufficiently large, then

$$\Delta\epsilon_{cv} = \epsilon_{cv}^p - \epsilon_{cv}^q \sim \mathcal{N}(\mu, \sigma^2), \quad (9)$$

where $\mu = \frac{m_{cv}}{m}(\varepsilon_g^p - \varepsilon_g^q)$, and $\sigma^2 = \frac{2m_{cv}}{m^2}[(\varepsilon_g^p)^2 + (\varepsilon_g^q)^2 - 2\rho_g\varepsilon_g^p\varepsilon_g^q]$.

Lemma 2 brings us closer to determining the probability that $\epsilon_{cv}^p > \epsilon_{cv}^q$ given that $\varepsilon_x^p > \varepsilon_x^q$. In particular, we obtain:

Theorem 2. *(Recovery error comparison) Let $\hat{\mathbf{x}}^p$ and $\hat{\mathbf{x}}^q$ be two recovered signals. If $\varepsilon_x^p \geq \varepsilon_x^q$, it holds with probability $\Phi(\lambda)$ that $\epsilon_{cv}^p \geq \epsilon_{cv}^q$, where λ is given by*

$$\frac{1}{\lambda^2} = \frac{2}{m_{cv}} \left[1 + 2(1 - \rho_g^2)\frac{\varepsilon_g^p\varepsilon_g^q}{(\varepsilon_g^p - \varepsilon_g^q)^2} \right], \quad (10)$$

and $\Phi(u)$ is the cumulative distribution function (CDF) of standard normal distribution,

$$\Phi(u) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{\frac{-t^2}{2}} dt. \quad (11)$$

Theorem 2 answers Prob. 2 by giving the probability that the comparison of CV residuals correctly reflect the comparison of the recovery errors. This probability increases with m_{cv}, ρ_g^2, and $\varepsilon_g^p/\varepsilon_g^q$. In other words, as the number of CV measurements increases, as recovered signals are more correlated, and as the recovery error difference increases, this probability also increases.

4. SPARSITY- AND NOISE- ROBUST OMP

This section provides our analysis on OMP-CV. Before stating our main theorem, we first define:

Definition 2. *(Ratio of unrecovered signal and noise): The ratio $\alpha^p \in \mathbb{R}$, defined as*

$$\alpha^p \triangleq \frac{\|\mathbf{x}_{T\backslash T^p}\|_2}{\sigma_n}, \quad (12)$$

measures to what extent the signal $\hat{\mathbf{x}}$ has not been recovered by $\hat{\mathbf{x}}^p$.

Next, we provide our main theorem for OMP-CV:

Theorem 3. *In OMP-CV, assume that the oracle optimum is $\hat{\mathbf{x}}^o$ and $T \subset T^o$. For any recovered signal $\hat{\mathbf{x}}^p$ other than $\hat{\mathbf{x}}^o$:*

- *if $T\backslash T^p \neq \emptyset$, then $\epsilon_{cv}^o < \epsilon_{cv}^p$ with probability $\Phi(\lambda)$, where*

$$\lambda \geq \sqrt{\frac{m_{cv}}{2}} \sqrt{1 - \mathrm{g}(\alpha^p)}; \quad (13)$$

- *if $T\backslash T^p = \emptyset$ and if $\hat{\mathbf{x}}^p$ is the OMP-CV output, then with probability greater than $\{1 - (d - k)[1 - \Phi(\lambda_0)]\}$ we have*

$$\varepsilon_g^p \leq C_1\varepsilon_g^o, \quad (14)$$

where $\mathrm{g}(\alpha^p)$ is roughly proportional to $1/(\alpha^p)^2$, λ_0 is a constant chosen to decide the probability with which (14) holds, and C_1 is only related to λ_0 and m_{cv}.

Theorem 3 supports the recovery performance of OMP-CV. It divides recovered signals into two categories. For $\hat{\mathbf{x}}^p$ with $T\backslash T^p \neq \emptyset$, the probability $[1 - \Phi(\lambda)]$ decays sharply as α^p increases. Since this is the probability that $\epsilon_{cv}^p < \epsilon_{cv}^o$, it is nearly impossible for such $\hat{\mathbf{x}}^p$ to be the OMP-CV output. For example, $[1 - \Phi(\lambda)]$ would be less than 0.5% with $\alpha^p = 1$, and further drop to 0.0063% with $\alpha^p = 2$. As a result, the OMP-CV output recovers all indices of the support set with overwhelming probability. Further, if the OMP-CV output recovers all indices of the support set, its recovery error can be bounded by ε_g^o with high probability showing that the recovery error of OMP-CV output is very close to that of the oracle optimum.

Remark 1. *Parameter details of Theorem 3 are:*

$$\mathrm{g}(\alpha^p) = \frac{\beta_1(\alpha^p)^2 + \beta_2}{\beta_1(\alpha^p)^2 + \beta_2 + \max((\alpha^p)^2 - \beta_3\alpha^p - \beta_4, 0)^2}$$
$$\approx \beta_1 / \left[(\alpha^p)^2 + \beta_1 \right],$$
$$C_1 = 2C_0 + 1 + 2\sqrt{C_0^2 + C_0},$$
$$C_0 \leq \beta_5\lambda_0^2/(m_{cv} - 2\lambda_0^2),$$

where betas are decided by RIP constant [4] of the sampling matrix. E.g., if $\delta_d < 0.1$, the values of betas are: $\beta_1 = 2.08$, $\beta_2 = \beta_3 = 0.03$, $\beta_4 = 0.02$, and $\beta_5 = 0.0376$. Further, if, e.g., $(d - k) = 100$, setting $m_{cv} = 48$ and $\lambda_0 = 4$, produces a numerical form of (14): with probability 99.7% we have

$$\varepsilon_g^p \leq 1.47\varepsilon_g^o \quad (15)$$

Apart from the performance analysis on OMP-CV, we also note that the recovery error of OMP-CV can be estimated using its CV residual by directly applying Theorem 1.
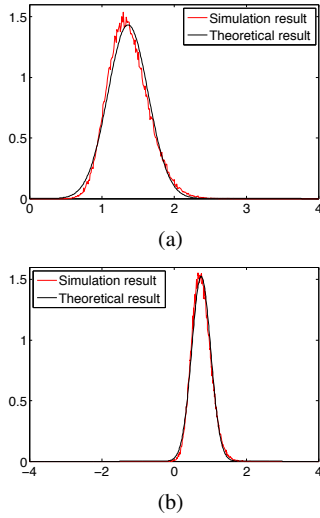
(a)

(b)

Fig. 1. Validation for Lemma 1 and Lemma 2. Fig.1(a) validates Lemma 1 by plotting the simulation result of probability distribution of ϵ_{cv}, while Fig.1(b) validates Lemma 2 by plotting that of $\Delta\epsilon_{\text{cv}}$ (red curve). The simulation results agree well with the theoretical results, which are plotted in black as reference. The Kullback-Leibler divergences of the theoretical results from the simulation results are 0.0152 and 0.0093 for Fig. 1(a) and Fig. 1(b) respectively.

5. NUMERICAL SIMULATION

This section gives our simulation results.[3] Section 5.1 simulates the probability distribution of random variables described in Lemma 1 and Lemma 2 to support both lemmas. Section 5.2 provides simulations on the performance of OMP-CV and OMP as a function of m_{cv}. The simulations demonstrate the recovery performance of OMP-CV. Throughout this section, Gaussian signals are used where non-zero entries are generated following the standard Gaussian distribution.

5.1. Validation for Lemma 1 and Lemma 2

As previously noted, CLT is used to approximate the probability distribution of ϵ_{cv} in Lemma 1 and $\Delta\epsilon_{\text{cv}}$ in Lemma 2. The simulations in this section attempt to validate this approximation. In both simulations, parameters are set as $N = 512$, $m = 96$, $m_{\text{cv}} = 48$, and $k = 50$. With recovered signals ($\hat{\mathbf{x}}$ for Lemma 1; $\hat{\mathbf{x}}^p$ and $\hat{\mathbf{x}}^q$ for Lemma 2) fixed, the random CV matrix along with its noise is realized 1E5 times and the probability distributions of random variables (ϵ_{cv} for Lemma 1; $\Delta\epsilon_{\text{cv}}$ for Lemma 2) are calculated. The experiment results, shown in Fig.1, indicate that the simulation results agree well with the theoretical prediction. This validates our approximation and supports both lemmas.

5.2. Recovery Performance of OMP-CV with Variation of CV Measurements Number

Given a fixed total number of measurements M, there is a tradeoff between m and m_{cv} [17]. On one hand, increasing m will reduce
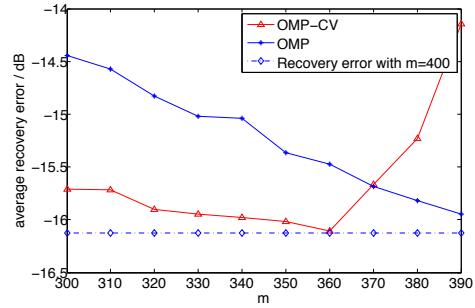
---

[3]The code for simulation is available at
http://gu.ee.tsinghua.edu.cn/publications#zj1

Fig. 2. The trade Off between m and m_{cv}. Parameters are fixed as $N = 1000$, $k = 50$, $\sigma_n^2 = 0.1$, and $M = 400$. m_{cv} varies from 100 to 10 with step -10 while $m = M - m_{\text{cv}}$. OMP-CV outperforms OMP except when m_{cv} is too small. In addition, using same number of measurements, OMP-CV has recovery performance similar to OMP (recovery error with $m = 400$) with parameters appropriately set even though the prior knowledge is required for OMP.

the reconstruction error. On the other hand, increasing m_{cv} will improve the CV estimate, and thus make the OMP-CV output closer to the oracle optimum. This simulation empirically investigates the recovery performance of OMP-CV as m_{cv} varies.

In this experiment, we set $N = 1000$, $k = 50$, $\sigma_n^2 = 0.1$, and $M = 400$. m_{cv} varies from 100 to 10 with step -10 and we have $m = M - m_{\text{cv}}$. For OMP-CV, m measurements are used for reconstruction, while m_{cv} measurements are used for CV. For OMP, m measurements are used for reconstruction and the termination is based on residual with the accurate noise level given. In addition, the recovery performance, where all M measurements are used for reconstruction using OMP, is given for comparison. We average 1000 repetitions for experiments of each parameter setting.

The experiment result, plotted in Fig.2, shows the best performance of OMP-CV lies in the region where m_{cv} is neither too small nor too large. OMP-CV outperforms OMP except when m_{cv} is very small, indicating that CV-based termination is better than residual-based termination, even if the latter uses an exact noise level. Additionally, note that with the same number of measurements at hand (M measurements for both OMP-CV and OMP), OMP-CV can achieve recovery performance similar to OMP with parameters appropriately set, even though prior knowledge is required for OMP. In this sense, OMP-CV outperforms OMP.

6. CONCLUSION

This paper presents a theoretical study of CV in compressive sensing, providing analysis of general CV problems as well as analysis of the OMP-CV algorithm. As a highly practical algorithm, OMP-CV could reconstruct the signal without prior knowledge such as sparsity or noise level; its performance is supported in this paper both theoretically and empirically. Additionally, our results on general CV problems could also be used to apply CV to other CS-based reconstruction algorithms. CV sacrifices a small amount of measurements to estimate the reconstruction error. In a nutshell, this technique makes it possible for greedy algorithms to reconstruct the signal without prior knowledge like the sparsity or noise level. In future work, we would like to extend our analysis on CV by studying the use of CV in other greedy sparse recovery algorithms.

7. REFERENCES

[1] Jinye Zhang, Laming Chen, Petros T. Boufounos, and Yuantao Gu, "Cross validation in compressive sensing and its application on omp-cv algorithm," submitted to *IEEE* for possible publication, available at http://gu.ee.tsinghua.edu.cn/publications#zj2.

[2] David L Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] Emmanuel J Candès, "Compressive sampling," in *Proceedings oh the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, 2006, pp. 1433–1452.

[4] Emmanuel J Candes and Terence Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, 2005.

[5] Emmanuel Candes, Mark Rudelson, Terence Tao, and Roman Vershynin, "Error correction via linear programming," in *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*. IEEE, 2005, pp. 668–681.

[6] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[7] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.

[8] Deanna Needell and Roman Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[9] David L Donoho, Yaakov Tsaig, Iddo Drori, and J-L Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1094–1121, 2012.

[10] Wei Dai and Olgica Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, 2009.

[11] Deanna Needell and Joel A Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[12] Pierre A Devijver and Josef Kittler, *Pattern recognition: A statistical approach*, Prentice/Hall International Englewood Cliffs, NJ, 1982.

[13] Seymour Geisser, *Predictive interference: an introduction*, vol. 55, CRC Press, 1993.

[14] Ron Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, 1995, vol. 14, pp. 1137–1145.

[15] Richard R Picard and R Dennis Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 575–583, 1984.

[16] Jun Shao, "Linear model selection by cross-validation," *Journal of the American statistical Association*, vol. 88, no. 422, pp. 486–494, 1993.

[17] Petros Boufounos, Marco F Duarte, and Richard G Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*. IEEE, 2007, pp. 299–303.

[18] Rachel Ward, "Compressed sensing with cross validation," *Information Theory, IEEE Transactions on*, vol. 55, no. 12, pp. 5773–5782, 2009.