Sparse signals estimation for adaptive sampling

Andrey Ordin [1], [*]

[1]*Lomonosov Moscow State University, Moscow, 119991*

Abstract This paper presents an estimation procedure for sparse signals in adaptive setting. We show that when the pure signal is strong enough, the value of loss function is asymptotically the same as for an optimal estimator up to a constant multiplier.

1. Introduction

The problem of sparse signal estimation is well studied, and optimal results have been established for various models (see, e.g., [1] and references therein). However, most existing approaches to the estimation assume non-adaptive sampling process. Adaptive sampling, on the other hand, leads to significant performance gains as well as sharper estimates of the signal. In this paper we propose a procedure for pure signal estimation under the sequential sampling framework introduced in [3].

Consider the following classical signal model

$$y_i = \mu_i + \varepsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where ε_i are independent standard normal random variables, and the vector $\mu = (\mu_1, \ldots, \mu_n)$ has most of its coordinates equal to zero. The problems of

---
[*]Correspondence to: Lomonosov Moscow State University, Moscow, 119991 E-mail: andrey.ordin@yahoo.com.

identification of non-zero components as well as the issues of their estimation from the observations $y_1, \ldots, y_n$ find applications in genetic microarray data analysis [6], astronomical surveying [5], and many other fields of science (see, e.g., [7]). A widespread approach to these problems is thresholding. The idea is to assume μ_i to be zero if the value of the corresponding observation y_i does not exceed some threshold, which may be chosen to depend on the whole vector y (e.g., [1]).

In [3], the authors consider a generalization of (1) that allows for sequential sampling procedures. In the proposed model one may construct multi-step algorithms such that the estimations carried out at the step j may depend on the results of all the preceding steps. Additionally, the authors introduce an algorithm called *Distilled Sensing* which helps to identify almost all of the zero and non-zero components correctly. Several results on the optimal estimation under adaptive setting for the case of two experiments have been also established in recent work [4].

While most existing papers on the topic are devoted to identification of non-zero components of the signal, we focus on its estimation. In the setup with an unlimited number of experiments, we suggest a way to estimate a sparse vector μ by slightly modifying the *Distilled Sensing* algorithm. We investigate the precision of the resulting estimates and derive sufficient conditions for their asymptotic optimality.

2. Adaptive scheme

We consider the model introduced in [3]. In contrast to the classical setup (1), there are $k \in \mathbb{N}$ noised observations of the signal $\mu \in \mathbb{R}^n$. To each observation $j = 1, \ldots, k$ there correspond some non-negative amount of energy $\mathcal{E}_j$ and a random set $I_j \subseteq \{1, \ldots, n\}$. The observations have the form

$$y_{ij} = \sqrt{\phi_{ij}}\mu_i + \varepsilon_{ij}, \quad j = 1, \ldots, k \quad i = 1, \ldots, n \qquad (2)$$

where ε_{ij} are independent standard normal random variables and

$$\phi_{ij} = \begin{cases} \frac{\mathcal{E}_j}{|I_j|} & \text{for } i \in I_j, \\ 0 & \text{otherwise.} \end{cases} \qquad (3)$$

Furthermore, the total amount of energy is limited, namely $\mathcal{E}_1 + \ldots + \mathcal{E}_k = n$. Thus, model (2) is reduced to (1) when $k = 1$ and $I_1 = \{1, \ldots, n\}$.

Denote by $S_n(\mu) = \{i : 1 \le i \le n, \mu_i \ne 0\}$ the set of all non-zero components of μ. We focus on the case of sparse signals when $S_n(\mu)$ constitutes a very small share of n; throughout this paper it is assumed that $S_n(\mu) \le n^{1-\beta}$ for all n large enough and some β from $(0, 1/3)$. Since it is not known beforehand which components of μ are not zero, the value of loss function (in the sense of (7) below)

in model (1) exceeds $n^{1-\beta}$. Moreover, reasonably precise estimations of the set $S_n(\mu)$ require additional assumptions on the vector μ (see, e.g., [3]). On the other hand, under adaptive sampling these assumptions may be significantly relaxed. We will restrict our attention to the signals μ from the set

$$\Theta_n(r) = \{\mu \in \mathbb{R}_+^n : |S_n(\mu)| \leq n^{1-\beta}, \ \mu_i \geq r\sqrt{\log n} \text{ for all } i \in S_n(\mu)\} \quad (4)$$

where $r > 0$.

In our approach to pure signal estimation we choose the energy levels slightly differently than it is done in [3]. Fix $\delta \in (0, 1)$ and set

$$\mathcal{E}_j = \begin{cases} \frac{\delta n}{4}\left(1 - \frac{\delta}{2}\right)^{j-1}, & j = 1, \ldots, k-2 \\ \frac{n}{2}\left(1 - \frac{\delta}{2}\right)^{k-2}, & j = k-1 \\ \frac{n}{2}, & j = k. \end{cases} \quad (5)$$

We follow the *Distilled Sensing* algorithm (DS) introduced in [3]. It is assumed that $j = 1, \ldots, k$ are sequential experiments, and ϕ_{ij} corresponds to the energy spent on suppressing the noise for the i-th component on the j-th step. For the given $\mathcal{E}_j$ the sets I_j are determined as follows.

Distilled Sensing
Step 0: Set $I_1 = \{1, \ldots, n\}$.
Step j: Set $I_{j+1} = I_j \cap \{i : y_{ij} > 0\}$.
　　　　　Determine ϕ_{ij} as in (3) above, $i = 1, \ldots, n$.
Result: Obtain the observations y_{ij} and a finite sequence of sets I_j.

Note that the sets $I_2, \ldots, I_{k+1}$ are random while I_1 is not. Correspondingly, for $i = 1, \ldots, n$, coefficients ϕ_{i1} are non-random but coefficients ϕ_{ij} with $j \geq 2$ are.

The authors of DS use I_{k+1} as an estimate for the set $S_n(\mu)$ of non-zero components for μ. In this paper we estimate the vector μ itself. Namely, as an estimate of μ take the vector $\hat{\mu}$ with components

$$\hat{\mu}_i = \begin{cases} 0, & \phi_{ik} = 0, \\ \frac{y_{ik}}{\sqrt{\phi_{ik}}}, & \phi_{ik} \neq 0, \end{cases} \quad (6)$$

where $i = 1, \ldots, n$. Now consider a loss function

$$R(\hat{\mu}, r) = \sup_{\mu \in \Theta_n(r)} \mathsf{E}\|\hat{\mu} - \mu\|^2, \quad (7)$$

where $\|z\|^2 = z_1^2 + \ldots + z_n^2$ for $z \in \mathbb{R}^n$ and $\Theta_n(r)$ is as in (4).

Theorem 2.1
Let $\delta \in (0, 1)$ and $\beta \in (0, 1/3)$. If the number of experiments k is large enough,

there exists $K > 0$ such that for $r > \sqrt{10\beta/\delta}$ and n large enough the following inequality holds

$$R(\hat{\mu}, r) < K n^{1-2\beta}.$$

It should be noted that in non-adaptive settings of (1) an equivalent loss function takes much larger values. In particular, asymptotic losses are of order $C(\beta)n^{1-\beta}\log n$ when μ has approximately $n^{1-\beta}$ non-zero components (see [6]). Furthermore, the losses of the order $n^{1-2\beta}$ are in a certain sense asymptotically optimal when $|S_n(\mu)| \sim n^{1-\beta}$ as $n \to \infty$. This optimality is studied more closely in Section 3, and the proof of Theorem 2.1 is provided in Section 4.

3. Optimal estimation

Proposition 3.1
Assume that (2) is satisfied and that the set of non-zero coordinates $S_n(\mu)$ is known beforehand. Whatever the choices of $k(n)$, $\mathcal{E}_j$, and I_j are, there exists no linear estimator $\hat{\mu}(y_{ij})$ for μ such that expected losses for it are less than $|S_n(\mu)|^2/n$ for each $n \in \mathbb{N}$.

Proof
Suppose that some k, $\mathcal{E}_j$, and I_j have been chosen. Their choice determines the coefficients ϕ_{ij} according to (3). Fix $i \in \{1, \ldots, n\}$. Since the observations corresponding to μ_i are of the form $y_{ij} = \sqrt{\phi_{ij}}\mu_i + \varepsilon_{ij}$, $j = 1, \ldots, k$, a standard argument shows that

$$\mathsf{E}(\hat{\mu}_i - \mu_i)^2 \geq \left(\sum_{j=1, \phi_{ij} \neq 0}^{k} \phi_{ij} \right)^{-1}.$$

Since all zero coordinates are known in advance, for $i \notin S_n(\mu)$ the best estimate is $\hat{\mu}_i = 0$. Thus one has

$$\mathsf{E}(\hat{\mu} - \mu)^2 \geq \sum_{i \in S_n(\mu)} \left(\sum_{1 \leq j \leq k, \phi_{ij} \neq 0} \phi_{ij} \right)^{-1}$$

$$\geq \inf\left\{ \sum_{i \in S_n(\mu)} \left(\sum_{1 \leq j \leq k, \psi_{ij} \neq 0} \psi_{ij} \right)^{-1} : \psi_{ij} \geq 0, \sum_{i \in S_n(\mu)} \sum_{j=1}^{j=k} \psi_{ij} \leq n \right\}.$$

The infimum on the right-hand side can be evaluated by solving the underlying optimization problem with Lagrange multipliers method. Its solution implies that

$$\mathsf{E}(\hat{\mu} - \mu)^2 \geq |S_n(\mu)|^2/n. \quad \square$$

4. Loss under distilled sensing

Proof of Theorem 2.1
As described in the section 2, for $k \in \mathbb{N}$ the sets I_j are chosen according to the DS algorithm, and the estimate $\hat{\mu}$ is defined by (6). The loss function (7) can be rewritten as follows.

$$R(\hat{\mu}, r) = \mathsf{E} \left(\sum_{i \in I_k \setminus S_n(\mu)} (\hat{\mu}_i - \mu_i)^2 + \sum_{i \in I_k \cap S_n(\mu)} (\hat{\mu}_i - \mu_i)^2 + \sum_{i \in S_n(\mu) \setminus I_k} (\hat{\mu}_i - \mu_i)^2 \right).$$

Here, the first sum is taken over the coordinates of μ falsely considered to be non-zero, the second sum is taken over correctly identified non-zero coordinates, and the third sum is taken over the coordinates misidentified as zeroes. We denote these loss components R_1, R_2, and R_3 respectively and treat them separately.
Estimation of R_1.

Let $s_j = |\{I_j \setminus S(\mu)\}|$ for $j = 1, \ldots, n$. Denote $\lceil x \rceil = \min\{m \in \mathbb{N} : m \geq x\}$. The following result holds.

Proposition 4.1
For a fixed ε_0 from $(0, 1/2)$ set $k = 2 + \lceil 2\beta \log_{\frac{2}{1+2\varepsilon_0}} n \rceil$. Then for all n large enough

$$\mathsf{P}\left(s_k \leq n^{1-2\beta}\right) \geq 1 - \exp\{-n^{1-2\beta}\varepsilon_0^2\}.$$

Proof
Note that $\mathsf{P}(\varepsilon_{ij} > 0) = 1/2$. By the Hoeffding inequality with probability 1 we get

$$\mathsf{E}\left(\mathbb{I}\left\{ \frac{1}{2} - \varepsilon_0 \leq \frac{s_{j+1}}{s_j} \leq \frac{1}{2} + \varepsilon_0 \right\} \Big| s_j \right) \geq 1 - 2\exp\{-2s_j\varepsilon_0^2\},$$

where $\mathbb{I}\{A\}$ is the indicator of an event A.

Let $A_{j+1} = \{s_{j+1} \leq (\varepsilon_0 + 1/2)s_j\}$ and $A = \bigcap_{j=1}^{k-1} A_{j+1}$. Denote by $\bar{A}$ the complement of A. Note that by the choice of k one has $\mathsf{P}(A \cap \{s_k > n^{1-2\beta}\}) = 0$. Moreover, the sequence $\{s_j\}_{j=1}^{k}$ is decreasing, therefore, $\{s_k > n^{1-2\beta}\} \subseteq \{s_j > n^{1-2\beta}\}$ for $j = 1, \ldots, k-1$. Thus for n large enough we have

$$\mathsf{P}(s_k > n^{1-2\beta}) = \mathsf{P}(\{s_k > n^{1-2\beta}\} \cap \bar{A}) \leq \sum_{j=1}^{k-1} \mathsf{P}(\{s_k > n^{1-2\beta}\} \cap \bar{A}_{j+1})$$

$$\leq 2k \exp\{-2n^{1-2\beta}\varepsilon_0^2\} < \exp\{-n^{1-2\beta}\varepsilon_0^2\},$$

which completes the proof. $\qquad\square$

Fix some $i \in I_k \setminus S_n(\mu)$. Note that $\phi_{ik} \geq \mathcal{E}_k/n = 1/2$. Hence for such i there holds the inequality $\mathsf{E}(\hat{\mu}_i - \mu_i)^2 \leq 2$. For n large enough we have

$$R_1 = \mathsf{E} \sum_{i \in I_k \setminus S_n(\mu)} (\hat{\mu}_i - \mu_i)^2 \leq \mathsf{E} 2 s_k \leq 3n^{1-2\beta}.$$

Estimation of R_2.

By the definition of s_k we have $|I_k| \leq |S_n(\mu)| + s_k$. Let k and ε_0 be as in Proposition 4.1. One has

$$\mathsf{P}\left(|I_k| > |S_n(\mu)| + n^{1-2\beta}\right) \leq \mathsf{P}\left(s_k > n^{1-2\beta}\right) \leq \exp\{-n^{1-2\beta}\varepsilon_0^2\}.$$

For a fixed $i \in I_k \cap S_n(\mu)$ one has

$$\mathsf{P}(\phi_{ik} \leq n^\beta/2) = \mathsf{P}(|I_k| \geq n^{1-\beta}) \leq \exp\{-n^{1-2\beta}\varepsilon_0^2\}.$$

Moreover, $\phi_{ik} \geq \mathcal{E}_k/n = 1/2$. Consequently,

$$\mathsf{E}(\hat{\mu}_i - \mu_i)^2 = \mathsf{E}\mathsf{E}\left((\hat{\mu}_i - \mu_i)^2 \big| \phi_{ik}\right) = \mathsf{E}1/\phi_{ik}$$

$$\leq \mathsf{E}\mathbb{I}\{\phi_{ik} > n^\beta/2\}/\phi_{ik} + 2\mathsf{E}\mathbb{I}\{\phi_{ik} \leq n^\beta/2\} \leq 2n^{-\beta}.$$

Therefore,

$$R_2 = \sum_{j \in I_k \cap S_n(\mu)} \mathsf{E}(\hat{\mu}_j - \mu_j)^2 \leq \sum_{j \in I_k \cap S_n(\mu)} 2n^{-\beta} \leq 2n^{1-2\beta}.$$

Estimation of R_3.

Let $\Phi(x)$ and $\varphi(x)$ be cumulative distribution function and density of the standard normal distribution respectively. Fix i from $S_n(\mu)$. For $j \geq 2$ observe that

$$\mathsf{P}(i \in I_j) = \mathsf{P}(i \in I_j, i \in I_{j-1}) = \mathsf{P}(i \in I_{j-1})\mathsf{P}(i \in I_j | i \in I_{j-1}),$$

and, consequently,

$$\mathsf{P}(i \in I_k) = \mathsf{P}(i \in I_2) \prod_{j=3}^{k} \mathsf{P}(i \in I_j | i \in I_{j-1}).$$

Since ϕ_{i1} for $i \in S_n(\mu)$ are fixed, $\mathsf{P}(i \in I_2) = \mathsf{P}(\varepsilon_{i1} > -\mu_i\phi_{i1}) = \Phi(\mu_i/2)$. Furthermore, for $j = 3, \ldots, k$ one has

$$\mathsf{P}(i \in I_j | i \in I_{j-1}) = \mathsf{P}(\varepsilon_{i,j-1} > -\mu_i\sqrt{\phi_{i,j-1}}|\phi_{i,j-1} > 0). \qquad (8)$$

To estimate the expression on the right-hand side we establish the following result.
Proposition 4.2
For every $\delta \in (0,1)$ and $\varepsilon_0 \in (0, (1-\delta)/2)$ one has

$$\mathsf{P}\left(s_j \leq n\left(\frac{1+2\varepsilon_0}{2}\right)^j \text{ for all } j = 1,\dots,k\right) \geq 1 - \exp\{-n^{1-2\beta}\varepsilon_0^2/2\}.$$

Proof
Set $A_j = \{s_j \leq n\left((1+2\varepsilon_0)/2\right)^j\}$ for $j = 1,\dots,k$. As it follows from the proof of Proposition 4.1,

$$\mathsf{P}(A_j) \geq 1 - \exp\{-n^{1-2\beta}\varepsilon_0^2\}.$$

Since $\mathsf{P}\left(\bigcap_{l=1}^{j} A_l\right) \geq \mathsf{P}(A_j) - \sum_{l=1}^{j-1}\mathsf{P}(\bar{A}_l)$, we have

$$\mathsf{P}\left(\bigcap_{j=1}^{k} A_j\right) \geq 1 - \exp\{-n^{1-2\beta}\varepsilon_0^2\} - (k-1)\exp\{-n^{1-2\beta}\varepsilon_0^2\}.$$

For all n large enough one obtains the estimate

$$\mathsf{P}\left(\bigcap_{j=1}^{k} A_j\right) \geq 1 - \exp\{-n^{1-2\beta}\varepsilon_0^2/2\}.$$

$\square$

Fix some $i \in S_n(\mu)$. Recall that $\mathcal{E}_j \geq \frac{\delta n}{4}\left(1 - \frac{\delta}{2}\right)^{j-1}$ for $j = 1,\dots,k$. According to the algorithm procedure, either $\phi_{ij} \geq \frac{\mathcal{E}_j}{n^{1-\beta}+s_j}$ or $\phi_{ij} = 0$. Using the above estimate for s_j one obtains

$$\mathsf{P}(\phi_{ij} > \delta/4 | \phi_{ij} > 0) \geq 1 - \exp\{-n^{1-2\beta}\varepsilon_0^2/2\}.$$

Set $A_j = \{\phi_{ij} > \delta/4\}$. Relation (8) yields

$$\mathsf{P}(i \in I_j | i \in I_{j-1}) \geq \mathsf{P}\left(A_{j-1} \cap \left\{\varepsilon_{i,j-1} > -\mu_i\sqrt{\phi_{i,j-1}}\right\}\Big|\phi_{i,j-1} > 0\right)$$

$$\geq \mathsf{P}\left(\varepsilon_{i,j-1} > -\mu_i\sqrt{\delta/4}\Big|\phi_{i,j-1} > 0\right) - \mathsf{P}\left(\bar{A}_{j-1}\Big|\phi_{i,j-1} > 0\right)$$

$$\geq \Phi(\mu_i\sqrt{\delta/4}) - \exp\{-n^{1-2\beta}\varepsilon_0^2/2\}. \qquad (9)$$

Therefore,

$$\mathsf{P}(i \in I_k) \geq \Phi(\mu_i/2)\prod_{j=3}^{k}\left(\Phi(\mu_i\sqrt{\delta/4}) - \exp\{-n^{1-2\beta}\varepsilon_0^2/2\}\right).$$

Using this estimate and the fact that $1 - \Phi(t) < \varphi(t)/t$ when $t > 0$, for all n large enough we get

$$\mathsf{P}(i \in I_k) \geq \left(1 - \frac{\varphi(\mu_i/2)}{\mu_i/2}\right) \prod_{j=3}^{k} \left(1 - \frac{\varphi(\mu_i\sqrt{\delta/4})}{\mu_i\sqrt{\delta/4}} - \exp\{-n^{1-2\beta}\varepsilon_0^2/2\}\right).$$

At first, consider the case when $\mu_i^2\delta/4 < n^{1-2\beta}\varepsilon_0^2/2$. For such μ_i and n large enough we derive the following estimate:

$$\mathsf{P}(i \in I_k) \geq \left(1 - \frac{\varphi(\mu_i/2)}{\mu_i/2}\right) \prod_{j=3}^{k} \left(1 - \varphi(\mu_i\sqrt{\delta}/4)\right) \geq \left(1 - \varphi(\mu_i\sqrt{\delta}/4)\right)^{k-1}.$$

Since $\mu_i \geq r\sqrt{\log n}$, for i from $S_n(\mu)$ and all n large enough, the expected contribution to the losses R_3 does not exceed

$$(1 - \mathsf{P}(i \in I_k))\,\mu_i^2 \leq (k-1)\mu_i^2\varphi(\mu_i\sqrt{\delta}/4) \leq n^{-r^2\delta/10} \leq n^{-\beta}. \qquad (10)$$

When $\mu_i^2\delta/4 \geq n^{1-2\beta}\varepsilon_0^2/2$, a different approach is required. Note that, for $j = 1, \ldots, k$, we have $\mathcal{E}_j \geq n^{1-\beta}$. Consequently, it follows that

$$\mathsf{P}(\phi_{ij} > n^{-\beta}|\phi_{ij} > 0) \geq \mathsf{P}(\phi_{ij} > \mathcal{E}_j/n|\phi_{ij} > 0) = 1.$$

Starting from the equation (8) and repeating the argument leading to (9), for $j = 3, \ldots, k$ we obtain

$$\mathsf{P}(i \in I_j|i \in I_{j-1}) \geq \Phi(\mu_i n^{-\beta/2})$$

and

$$\mathsf{P}(i \in I_k) \geq \Phi(\mu_i/2) \prod_{j=3}^{k} \Phi(\mu_i n^{-\beta/2}) \geq \left(\Phi(\mu_i n^{-\beta/2})\right)^{k-1} \geq \left(1 - \frac{\varphi(\mu_i n^{-\beta/2})}{\mu_i n^{-\beta/2}}\right)^{k-1}.$$

For such i with $\mu_i^2 \geq 2n^{1-2\beta}\varepsilon_0^2/\delta$ one has

$$\mathsf{E}(\hat{\mu}_i - \mu_i)^2 = (1 - \mathsf{P}(i \in I_k))\,\mu_i^2 \leq \mu_i\varphi(\mu_i^2 n^{-\beta}) \leq \exp\{-cn^{1-3\beta}\} \qquad (11)$$

where $c = c(\delta, \varepsilon_0) > 0$. From (10) and (11) it follows that

$$R_3 = \mathsf{E} \sum_{i \in S_n(\mu)\backslash I_k} (\hat{\mu}_i - \mu_i)^2 \leq \sum_{i \in S_n(\mu)} \mathsf{P}(i \notin I_k)\mu_i^2 \leq |S_n(\mu)|n^{-\beta} \leq n^{1-2\beta}.$$

Combining the results of this section, we get

$$R = R_1 + R_2 + R_3 \leq 3n^{1-2\beta} + 2n^{1-2\beta} + n^{1-2\beta} = 6n^{1-2\beta}$$

which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5. Concluding remarks

In this paper we suggest an estimation procedure for sparse signals in adaptive setting based on the *Distilled Sensing* algorithm introduced in [3]. In a sense, our results are optimal when the set of non-zero components of pure signal is of the order $n^{1-\beta}$ with some $\beta \in (0, 1/3)$. For this case we demonstrate that the asymptotic behavior of loss function is the same as for an optimal estimator up to a constant multiplier. Under general assumptions such performance cannot be achieved in non-adaptive setting. Further research could cover the cases of weaker signals when μ_i are not bounded from below.

REFERENCES

1. F. Abramovich, Y. Benjamini, D. L. Donoho, I. M. Johnstone. Adapting to unknown sparsity by controlling false discovery rate. *The Annals of Statistics*, 34, (2006), 584-653.
2. T. T. Cai, X. J. Jeng, J. Jin. Optimal detection of heterogeneous and heteroscedastic mixtures. *Journal of the Royal Statistical Society: Series B*, 73(5), (2011), 629-662.
3. J. Haupt, R. Castro, R. Nowak. Distilled Sensing: Selective Sampling for Sparse Signal Recovery, (2009), *International Conference on Artificial Intelligence and Statistics*.
4. A. O. Hero, III, D. Wei. Performance guarantees for adaptive estimation of sparse signals, (2013), arxiv preprint arXiv:1311.6360v1.
5. A. M. Hopkins. A new source detection algorithm using the false-discovery rate, *Astron. J.*, 123, (2002), 1086-1094.
6. Y. Pawitan, S. Michiels, S. Koscielny, A. Gusnanto, A. Ploner. False discovery rate, sensitivity, and sample size for microarray studies. *Bioinformatics*, Vol. 21, (2005), 3017-3024.
7. J. L. Starck, F. Murtagh, J. M. Fadili. *Sparse Image and Signal Processing*, (2010), Cambridge University Press.