Abstract—In this paper it is shown that a modification of the Elastic Net algorithm (MEN) exhibits near ideal behavior in the following sense: Suppose the input to the algorithm is a vector of known sparsity index but unknown locations for the nonzero components. Then the output error of the algorithm is bounded by a universal constant times the error achieved by an oracle that knows not just the sparsity index but also the locations of the nonzero components. This result generalizes an earlier result of Candès and Plan on the near ideal behavior of the LASSO algorithm.

What follows is an extended abstract of the paper.

A well-known result of Candès and Tao [1] states the following: Suppose $x \in \mathbb{R}^n$ but is $k$-sparse; thus $x$ has at most $k$ nonzero components, but the location of the nonzero components is not known. Suppose $A$ is an $m \times n$ matrix that satisfies the so-called Restricted Isometry Property (RIP) of order $2k$ with a coefficient $\delta_{2k} < \sqrt{2} - 1$. Then one can recover $x$ exactly by minimizing $\|z\|_1$ subject to $Az = y = Ax$. A later paper by Candès and Plan [2] studies the case where $y = Ax + \eta$ when $\|\eta\|_2 \leq \epsilon$. It is shown that minimizing $\|z\|_1$ subject to the condition that $\|y - Az\|_2 \leq \epsilon$ leads to “near-ideal behavior.” Specifically, the estimation error $\|z - x\|_2$ is bounded by a universal constant times the error achieved by an “oracle” that knows the location of the nonzero components of $x$.

The minimization of the $\ell_1$-norm is closely related to the so-called “LASSO” algorithm. LASSO is a special case of the so-called “Elastic Net” (EN) algorithm, that minimizing a convex combination of the $\ell_1$-norm and the square of the $\ell_2$-norm of the residuals. However, to date there are no results on the near-ideal behavior of the Elastic Net algorithm. Indeed it would be difficult to achieve this, as the quantity being minimized in the Elastic Net algorithm is not a norm. So in this paper we study a modified elastic net (MEN) algorithm wherein a convex combination of the $\ell_1$-norm and the $\ell_2$-norm (not the $\ell_2$-norm squared) is minimized. It is shown that the MEN algorithm also exhibits near ideal behavior. Specifically, if one were to minimize $(1 - \mu)\|z\|_1 + \mu\|z\|_2$, then under suitable conditions involving $\mu$ and $\delta_{2k}$, the estimation error $\|z - x\|_2$ is bounded by a universal constant times the error achieved by an “oracle” that knows the location of the nonzero components of $x$.

Definition 1: Let $\Sigma_k$ denote the set of all $x \in \mathbb{R}^n$ such that at most $k$ components of $x$ are nonzero. Suppose $A \in \mathbb{R}^{m \times n}$. Then we say that $A$ satisfies the Restricted Isometry Property (RIP) of order $k$ with constant $\delta_k$ if

\[
(1 - \delta_k)\|u\|_2^2 \leq \langle u, Au \rangle \leq (1 + \delta_k)\|u\|_2^2 \forall u \in \Sigma_k.
\]

Suppose the matrix $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$. Define

\[
\alpha = \frac{\sqrt{2}\delta_{2k}}{1 - \delta_{2k}}, \quad \beta = \frac{1}{1 - \delta_{2k}},
\]

\[
C_0 = 2\frac{1 + \alpha}{1 - \alpha} = 2\frac{1 + (\sqrt{2} - 1)\delta_{2k}}{1 - (\sqrt{2} + 1)\delta_{2k}}.
\]

The next result shows that the LASSO algorithm has near ideal behavior.

Theorem 1: (As stated in [3, Theorem 1.9]; compare also with [2, Theorem 1.4]) Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$, and that $y = Ax + \eta$ for some $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^n$ with $\|\eta\|_2 \leq \epsilon$. Let $B_\epsilon(y) := \{z \in \mathbb{R}^n : \|y - Az\|_2 \leq \epsilon\}$, and define

\[
\hat{x} = \arg\min_{z \in B_\epsilon(y)} \|z\|_1.
\]

Then

\[
\|\hat{x} - x\|_2 \leq C_0 \frac{\sigma_k(x, \cdot \|\cdot\|_1)}{\sqrt{k}} + C_2 \epsilon,
\]

where $C_0$ is as in (3) and

\[
C_2 = \frac{4\sqrt{1 + \delta_{2k}}}{(1 - \delta_{2k})(1 - \alpha)} = \frac{4\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}}.
\]

Theorem 2 below is the main result of the paper. It shows that the MEN algorithm also exhibits nearly ideal behavior. Moreover, the error bounds in Theorem 2 reduce to (5) when $\mu = 0$. Hence Theorem 2 is a true generalization of Theorem 1.

Theorem 2: Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k}$, and that $y = Ax + \eta$ for some $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^n$ with $\|\eta\|_2 \leq \epsilon$. Let $B_\epsilon(y) := \{z \in \mathbb{R}^n : \|y - Az\|_2 \leq \epsilon\}$, and define

\[
\hat{x}_{\text{MEN}} := \arg\min_{z \in B_\epsilon(y)} \|z\|_\mu.
\]

Define $\alpha$ as in (2), and suppose that $\delta_{2k} < \sqrt{2} - 1$, which guarantees that $\alpha < 1$. Define

\[
\gamma = \frac{1}{\sqrt{k}} \frac{\mu}{1 - \mu},
\]

and suppose $\mu$ and $\delta_{2k}$ together satisfy

\[
\gamma < \frac{1 - \alpha}{1 + \alpha} = \frac{1 - (\sqrt{2} + 1)\delta_{2k}}{1 + (\sqrt{2} - 1)\delta_{2k}}.
\]

Then

\[
\|\hat{x}_{\text{MEN}} - x\|_2 \leq C_0 \frac{\sigma_k(x, \cdot \|\cdot\|_1)}{\sqrt{k}} + C_2 \mu \epsilon,
\]
where

\[
C_{0,\mu} = \frac{2(1 + \alpha)}{1 - \alpha - \alpha \gamma},
\]

(11)

\[
C_{2,\mu} = \frac{4\sqrt{1 + \delta_{2k}}}{(1 - \delta_{2k})(1 - \alpha - \alpha \gamma)}.
\]

(12)

REFERENCES

