Past work studied this problem in various forms, and in this chapter we discuss some of what is now known. Results are in some ways parallel to those in the noiseless case. Specifically, we should discuss the uniqueness property – conditions under which a sufficiently sparse solution is known to be the global minimizer of \((P_0^\epsilon)\); practical pursuit techniques to approximate the solution of this problem; and equivalence – theoretical guarantees for their successful recovery of the desired solution. As we shall see next, however, the notions of uniqueness and equivalence no longer apply – they are replaced by the notion of stability.

5.2 Stability of the Sparsest Solution

Before turning to approximate the solution of \((P_0^\epsilon)\), we must answer a far more fundamental question: Suppose that a sparse vector \(x_0\) is multiplied by \(A\), and we obtain a noisy observation of this product, \(b = Ax_0 + e\) with \(\|b - Ax_0\|_2 \leq \epsilon\). Consider applying \((P_0^\epsilon)\) to obtain an approximation to \(x_0\), and getting a solution \(x_0^\epsilon\),

\[
x_0^\epsilon = \arg \min_x \|x\|_0 \text{ subject to } \|b - Ax\|_2 \leq \epsilon.
\]

How good shall this approximation be? How its accuracy is affected by the sparsity of \(x_0\)? These questions are the natural extension for the uniqueness property of sparse solutions we have discussed in the context of the \((P_0)\) problem in Chapter 2.

5.2.1 Uniqueness versus Stability – Gaining Intuition

As we show next, we can no longer claim uniqueness for \((P_0^\epsilon)\) in the general case. In order to demonstrate that, we present a simple experiment: \(A\) is chosen to be the two-ortho \([I, F]\) of size \(2 \times 4\), \(x_0 = [0 0 1 0]^T\). We generate a random noise \(e\) with pre-specified norm \(\|e\|_2\) and create the vector \(b = Ax_0 + e\). Thus,

\[
Ax_0 + e = \begin{bmatrix} 1 & 0 & 0.707 & 0.707 \\ 0 & 1 & 0.707 & -0.707 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 0.707 + e_1 \\ 0.707 + e_2 \end{bmatrix}
\]

Figure 5.1 presents the locations of \(Ax_0\), \(b\), the regions \(\{v | \|v - Ax_0\|_2 \leq \epsilon\}\), and \(\{v | \|v - b\|_2 \leq \epsilon\}\), with \(\epsilon = 0.2\). The last of those is the region of the image (i.e., after multiplication by \(A\)) of all the feasible solutions \(x\) to the problem \((P_0^\epsilon)\). Naturally, the vector \(x_0\) poses a feasible solution, while being very sparse. Indeed, it is an optimal solution to \((P_0^\epsilon)\), in the sense that no sparser solution exists (a sparser solution is only the null vector, and it is outside the feasible set). Could there be alternative feasible
solutions that are sparse? Figure 5.1 shows that solutions of the form $x_0 = [0 0 x 0]^T$ with some values of $x$ are also feasible, while having the same cardinality.

Figure 5.2 presents the same experiment, this time with a stronger noise, $\epsilon = 0.6$. This leads to a different scenario, where not only we have lost uniqueness with respect to the same support, but other supports with cardinality $\|x\|_0 = 1$ are possible, and in fact, even the null solution is included, implying that this is the optimal solution to $(P^\epsilon_0)$.

Here is a more formal way of explaining this. We shall denote $x_S$ and $A_S$ the portions of $x$ and $A$ that contain the support $S$ elements/columns, respectively. Suppose that $x$ is a sparse candidate solution to this problem over the support $S$, with $\|x\|_0 = |S|$, and it satisfies the constraint, $\|b - A_Sx_S\|_2 \leq \epsilon$.

If it so happens that $x_S$ is also the minimizer of the term $f_S(z) = \|b - A_Sz\|_2$, and $f_S(x_S^{opt}) = \epsilon$, we can propose no alternative solution over this support, since any perturbation around $x_S$ leads to an increase in this term and thus violation of the constraint. In terms of Figure 5.1, this case takes place when the closest point to $b$ on the green line is $Ax_0$, or put differently, if the distortion $e = b - A_Sx_S$ is orthogonal to the columns of $A_S$.\(^1\) In all other cases, the fact that min$_{z} f_S(z) < \epsilon$ implies an ability to perturb the so-called optimal solution $x_S$ in a way that preserves its feasibility and the support, and thus we get a set of solutions that are as good as $x$.

\(^1\) As the minimizer of $f_S(z) = \|b - A_Sz\|_2$, the vector $x_S$ should satisfy $A_S^T(b - A_Sx_S) = A_S^Te = 0$. 

---

Fig. 5.1 A 2D demonstration of the lack of uniqueness for the noisy case, with a relatively weak noise.
Furthermore, if some of the non-zero entries in $x$ are small enough, this perturbation may null them, leading to a sparser solution.

### 5.2.2 Theoretical Study of the Stability of \((P_0^c)\)

So, returning to the original question we have posed, instead of claiming uniqueness of a sparse solution, we replace this with a notion of stability – a claim that if a sufficiently sparse solution is found, then all alternative solutions necessarily resides very close to it. The following analysis, taken from the work by Donoho, Elad, and Temlyakov, leads to a stability claim of this sort.

We start by returning to the definition of the \textit{spark} and extending it by considering a relaxed notion of linear-dependency. In the noiseless case we considered two competing solutions $x_1$ and $x_2$ to the linear system $Ax = b$, and this led to the relation $A(x_1 - x_2) = Ad = 0$. This motivates a study of the sparsity of vectors $d$ in the null-space of $A$, which naturally leads to the definition of the \textit{spark}.

Following the same rationale, we should consider now two feasible solutions $x_1$ and $x_2$ to the requirement $\|Ax - b\|_2 \leq \epsilon$. Considering $b$ as the center of a sphere of radius $\epsilon$, both $Ax_1$ and $Ax_2$ reside in it or on its surface. Thus, the distance between

---

**Fig. 5.2** A 2D demonstration of the lack of uniqueness for the noisy case, as shown in Figure 5.1, but with a stronger noise, that permits alternative solutions with a different support.