7 Super-resolution radar imaging via convex optimization

Reinhard Heckel

Department of Electrical and Computer Engineering, Rice University

A radar system emits probing signals and records the reflections. Estimating the relative angles, delays, and Doppler shifts from the received signals allows to determine the locations and velocities of objects. However, due to practical constraints, the probing signals have finite bandwidth B, the received signals are observed over a finite time interval of length T only, and a radar typically has only one or a few transmit and receive antennas. These constraints fundamentally limit the resolution up to which objects can be distinguished. Specifically, a radar can not distinguish objects with delay and Doppler shifts much closer than $1/B$ and $1/T$, respectively, and a radar system with N_T transmit and N_R receive antennas cannot distinguish objects with angels closer than $1/(N_T N_R)$. As a consequence, the delay, Doppler, and angular resolution of standard radars is proportional to $1/B$ and $1/T$, and $1/(N_T N_R)$. In this chapter, we show that the *continuous* angle-delay-Doppler triplets and the corresponding attenuation factors can be resolved at much finer resolution, using ideas from compressive sensing. Specifically, provided the angle-delay-Doppler triplets are *separated* either by factors proportional to $1/(N_T N_R - 1)$ in angle, $1/B$ in delay, or $1/T$ in Doppler direction, they can be recovered a significantly smaller scale or higher resolution.

7.1 Introduction

A traditional Single-Input Single-Output (SISO) pulse-Doppler radar system transmits a probing signal and receives the reflections from objects with a single antenna. By estimating the induced delays and Doppler shifts the radar system can determine the relative distances and velocities of the objects. However— like for any imaging system—physics imposes a limit on how well objects can be resolved. The resolution of a radar system is determined by the bandwidth B of the probing signals and the time interval T over which the responses are observed. Specifically, the delay and Doppler resolution is proportional to $1/B$ and $1/T$, meaning that objects closer than that are essentially impossible to distinguish under noise. Since both B and T cannot be made arbitrarily large due to physical limitations, those two constraints fundamentally limit the resolution of a SISO radar system. In contrast to SISO radar systems, Mulitple-Input,

Multiple-Output (MIMO) radar systems [1, 2] use multiple antennas to transmit probing signals simultaneously and record the reflections from the objects with multiple receive antennas. A MIMO radar can thereby, in principle, resolve the relative angles in addition to the relative distances and velocities of objects with a single measurement. However, the angular resolution of a MIMO radar is $1/(N_T N_R)$, where N_T and N_R are the number of transmit and receive antennas, and is thus again limited by a physical constraint, namely the number of antennas (see Section 7.6 for a detailed argument on the resolution).

Even though objects that are simultaneously much closer than $1/(N_T N_R)$ in angle, $1/B$ in delay, and $1/T$ in Doppler direction, are impossible to distinguish for real world radar systems in general, it is still possible to determine the locations of the objects up to a much higher accuracy than the resolution limit of $(1/(N_T N_R), 1/B, 1/T)$. In this chapter, we discuss signal recovery techniques for building super-resolution radar systems that can achieve localization accuracy below the resolution limit. In more detail, we study the problem of recovering the *continuous* delays and Doppler shifts in a SISO radar system, and the problem of recovering the angles, delays and Doppler shifts in a MIMO radar system, in both cases from the responses to known and suitably selected probing signals. As we see later, those problems—termed the super-resolution radar and super-resolution MIMO radar problems—amount to recover a signal that is sparse in a *continuous* dictionary from linear measurements, and can thus be viewed as a generalization of the traditional compressive sensing problem.

If the objects may be assumed to lie on a sufficiently *coarse* grid, compressed sensing [3] based approaches provably recover the delay-Doppler pairs for SISO radar system [4, 5, 6], and the angle-delay-Doppler triplets for MIMO radar systems [7, 8]. However, to establish those results, the aforementioned papers assume that angles, delays, and Doppler shifts lie on a sufficiently *coarse* grid, specifically a grid with spacing $1/(N_T N_R), 1/B$, and $1/T$, in angle, delay, and Doppler direction, respectively (see Section 7.2.3). Since N_T, N_R, B, and T are physical problem parameters, they can in general not be made (arbitrarily) large in order to make the grid finer. In fact, the coarseness of the grid is required for the measurement matrix to be incoherent, therefore the aforementioned results cannot straightforwardly be extended to a grid with significantly finer spacing. In some special cases, however, off-the-grid recovery is possible with standard spectral estimation techniques. For example, for a single input antenna and either known and constant delays (see Section 7.2.4), or known and constant Doppler shifts, the super-resolution radar problem reduces to a standard 2D line spectral estimation problem [8, Sec. 5]. For these special cases, the object locations can be recovered—off the grid—with standard spectral estimation techniques such as Prony's method, MUSIC, and ESPRIT [9]. In general, however, the super-resolution radar problems cannot be reduced to the classical line spectral estimation problem. Therefore, traditional spectral estimation techniques are not directly applicable.

Recently, an alternative, convex optimization based approach to solve the clas-

sical *line spectral estimation* has been proposed that is much more generally applicable than traditional line spectral estimation techniques. Specifically, the paper [10] shows that the frequency parameters, which are the unknowns in the line spectral estimation problem, can be perfectly recovered by solving a convex total-variation norm minimization program, provided they are sufficiently separated. Related convex programs have been studied for compressive sensing off the grid [11], denoising [12], signal recovery from short-time Fourier measurements [13] and the SISO and MIMO super-resolution radar problems [14, 15], and the generalized line-spectral estimation problem [16]. The focus of this chapter is on explaining how this convex optimization based approach enables high resolution in radar. In particular we discuss the results in [14, 15], showing that a convex program recovers the *continuous* angles, delays, and Doppler shifts perfectly, provided that they are sufficiently separated. Furthermore, we show that a simple convex ℓ_1-minimization program recovers the angles and delay-Doppler shifts on an *arbitrarily fine* grid, again provided they are sufficiently separated. Finally, we provide numerical results demonstrating robustness to noise.

Outline:
The remainder of this chapter is organized as follows. In the first part we consider the SISO radar model. In more detail, Section 7.2 contains the radar model and formal problem statement, in Sections 7.3 and 7.3.2 we present the convex optimization based recovery approach and corresponding performance guarantees, and in Section 7.4 we show that ℓ_1-minimization recovers the locations on an arbitrarily *fine* grid. In Section 7.4.2, we provide numerical results demonstrating that the approach is robust to noise and in Section 7.5.2 we outline the proof of the main technical statements. In the second part, Section 7.6, we explain how the results for SISO radar can be extended to the MIMO case. We conclude in Section 7.7 with a discussion on challenges in applying those ideas in practice and current and future research directions.

7.2 Signal model and problem statement

A radar system with a single transmit and single receive antenna is typically modeled as a linear system. The response y recorded at the receive antenna, to a probing signal, x, emitted at the transmit antenna, is a weighted superposition of delayed and Doppler-shifted versions of the probing signal x:

$$y(t) = \iint s(\tau, \nu)x(t - \tau)e^{i2\pi\nu t}d\nu d\tau. \tag{7.1}$$

Here, s denotes the spreading function which describes the scene being sensed and τ and ν are the delays and Doppler shifts. Often, the moving objects are modeled by point scatterers. Mathematically, this means that the spreading function

specializes to

$$s(\tau, \nu) = \sum_{j=1}^{S} b_j \delta(\tau - \bar{\tau}_j) \delta(\nu - \bar{\nu}_j).$$

Here, b_j is the (complex-valued) attenuation factor associated with the delay-Doppler pair $(\bar{\tau}_j, \bar{\nu}_j)$. With the spreading function above, the input-output relation (7.1) reduces to

$$y(t) = \sum_{j=1}^{S} b_j x(t - \bar{\tau}_j) e^{i2\pi\bar{\nu}_j t}. \tag{7.2}$$

Thus, the received signal is a superposition of the reflections of the probing signal by the point scatterers. The relative distances and velocities of the S-many objects can be trivially obtained from the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$. In order to locate the objects, we therefore need to estimate the delay-Doppler pairs and the corresponding attenuation factors b_j from a single input-output measurement, i.e., from the response y to a known and suitably selected probing signal x. As we see later, the particular choice of the probing signal is crucial for good localization performance.

7.2.1   Band- and time-limitation and resolution

The probing signal x can be controlled by the system engineer and is known. However, due to practical and technological constraints, it must be band-limited and approximately time-limited. Also, again due to practical constraints, we can only observe the response y over a finite time interval. For concreteness, we assume that

i. we observe the response y over an interval of length T and that
ii. x has bandwidth B and is approximately supported on a time interval of length proportional to T.

The time- and band-limitation determines the "natural" resolution of the system, which is the accuracy up to which the delay-Doppler pairs can be identified. A standard pulse-Doppler radar that samples the received signal at its Nyqist rate, and performs digital matched filtering, estimates the parameters up to accuracy $1/B$ and $1/T$ in delay (τ) and Doppler (ν) directions, respectively, and therefore only identifies the delay-Doppler pairs up to the natural solution.

   From the input-output relation (7.2), it is evident that band- and approximate time-limitation of the input signal x implies that the response y is band- and approximately time-limited as well—provided that the delay-Doppler pairs are compactly supported. In radar, due to path loss and finite velocity of the objects in the scene this is indeed the case [17]. Throughout, we will therefore assume that the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ lie in the region

$$[-T/2, T/2] \times [-B/2, B/2].$$

This is not a restrictive assumption as this region can have area $BT \gg 1$, which is typically very large. In fact, for many applications, it is reasonable to assume that the delay-Doppler pairs lie in a region of area significantly smaller than one [18, 19, 20], an assumption often referred to as the linear system being "underspread". We do not make or require this assumption here.

By the $2WT$-Theorem [21], band- and approximate time-limitation of the response y implies that y is essentially characterized by on the order of BT coefficients. We therefore sample y in the interval $[-T/2, T/2]$ at rate $1/B$, so as to collect $L := BT$ samples, denoted by $y_p := y(p/B)$ (for simplicity we assume in the following that $L = BT$ is an odd integer). As detailed in [14, Sec. 5], those samples are given by

$$y_p = \sum_{j=1}^{S} b_j [\mathcal{F}_{\nu_j} \mathcal{T}_{\tau_j} \mathbf{x}]_p, \quad p = -N, ..., N, \quad N := \frac{L-1}{2}, \qquad (7.3)$$

where

$$[\mathcal{T}_\tau \mathbf{x}]_p := \frac{1}{L} \sum_{k=-N}^{N} \left[\left(\sum_{\ell=-N}^{N} x_\ell e^{-i2\pi \frac{\ell k}{L}} \right) e^{-i2\pi k\tau} \right] e^{i2\pi \frac{pk}{L}} \qquad (7.4)$$

and

$$[\mathcal{F}_\nu \mathbf{x}]_p := x_p e^{i2\pi p\nu}.$$

Here, we defined the time-shifts $\tau_j := \bar{\tau}_j / T$ and frequency-shifts $\nu_j := \bar{\nu}_j / B$. To avoid ambiguity, from here onwards we refer to $(\bar{\tau}_j, \bar{\nu}_j)$ as a delay-Doppler pair and to (τ_j, ν_j) as a time-frequency shift. From $(\bar{\tau}_j, \bar{\nu}_j) \in [-T/2, T/2] \times [-B/2, B/2]$ we have $(\tau_j, \nu_j) \in [-1/2, 1/2]^2$. Since $\mathcal{T}_\tau \mathbf{x}$ and $\mathcal{F}_\nu \mathbf{x}$ are 1-periodic in τ and ν, we assume in the remainder of the chapter without loss of generality that $(\tau_j, \nu_j) \in [0, 1]^2$. The operators $\mathcal{T}_\tau$ and $\mathcal{F}_\nu$ have an interesting interpretation as fractional time and frequency shift operators in $\mathbb{C}^L$. In fact, if the parameters τ and ν lie on a $(1/L, 1/L)$ grid, the operators $\mathcal{F}_\nu$ and $\mathcal{T}_\tau$ reduce to the "natural" time frequency shift operators in $\mathbb{C}^L$, i.e.,

$$[\mathcal{T}_\tau \mathbf{x}]_p = x_{p-\tau L} \text{ and } [\mathcal{F}_\nu \mathbf{x}]_p = x_p e^{i2\pi p\nu}.$$

The definition of a time shift in (7.4) as taking the Fourier transform, modulating the frequency, and taking the inverse Fourier transform is a very natural definition of a *continuous* time-shift $\tau_j \in [0, 1]$ of a *discrete* vector $\mathbf{x} = [x_0, \ldots, x_{L-1}]^T$.

Finally, note that to obtain the input-output relation (7.3) (see [14, Sec. 5]) from (7.2), a periodic sinc function is approximated with a finite sum of sinc functions (this is where partial periodization of x becomes relevant). Thus, if we take the probing signal to be essentially time-limited, then equality in (7.3) does not hold exactly. However, in [14, Sec. 5] it is shown that for a random probing signal, as considered in this chapter, the incurred relative ℓ_2-error decays as $1/\sqrt{L}$ and is therefore negligible for large L. It is confirmed numerically in the same paper that the approximation error made in this process is negligible. Moreover,

if we took x to be T-periodic, then the input-output relation (7.3) becomes exact, but at the cost of the probing signal x not being time-limited.

7.2.2　Formal problem statement

From the discussion in the previous section we conclude that identification of the objects under the constraints that the probing signal x is band-limited and the response y to the probing signal is observed over a finite time interval, reduces to the estimation of the triplets $\{(b_j, \tau_j, \nu_j)\}_{j=1}^{S}$ from the samples $\{y_p\}_{p=-N}^{N}$. Thus, in this chapter, we consider the problem of recovering those triples from the samples $\{y_p\}_{p=-N}^{N}$ in (7.3). We call this the super-resolution radar problem, as recovering the exact time-frequency shifts $\{(\tau_j, \nu_j)\}_{j=1}^{S}$ "breaks" the natural resolution limit of $(1/B, 1/T)$ achieved by a standard pulse-Doppler radar.

Alternatively, one can view the super-resolution radar problem as that of recovering a signal that is S-sparse in the continuous dictionary of time-frequency shifts of an L-periodic sequence x_ℓ. In order to see this, and to better understand the super-resolution radar problem, it is instructive to consider two special cases.

7.2.3　Time-frequency shifts on a grid

Suppose the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ lie on a $(\frac{1}{B}, \frac{1}{T})$-grid. As a consequence the time-frequency shifts (τ_j, ν_j) lie on a $(\frac{1}{L}, \frac{1}{L})$-grid, which in turn implies that $\tau_j L$ and $\nu_j L$ are integers in $\{0, \dots, L-1\}$. Thus, the super-resolution radar problem reduces to a sparse signal recovery problem with a Gabor measurement matrix. To see this, note that under the aforementioned assumption, the input-output relation (7.3) reduces to

$$y_p = \sum_{j=1}^{S} b_j x_{p-\tau_j L} e^{i2\pi \frac{(\nu_j L)p}{L}}, \quad p = -N, \dots, N.$$

Writing this equation in vector-matrix form gives

$$\mathbf{y} = \mathbf{G_x} \mathbf{b}.$$

Here, the vector $\mathbf{y}$ contains as entries the samples y_p, $\mathbf{G_x} \in \mathbb{C}^{L \times L^2}$ is the Gabor matrix with window $\mathbf{x}$, defined by

$$[\mathbf{G_x}]_{p,(k,\ell)} := x_{p-\ell} e^{i2\pi \frac{kp}{L}}, \quad k, \ell, p = -N, \dots, N, \tag{7.5}$$

and $\mathbf{b} \in \mathbb{C}^{L^2}$ is a sparse vector with the j-th non-zero entry given by b_j and indexed by $(\tau_j L, \nu_j L)$.

Thus, recovery of the triplets $\{(b_j, \tau_j, \nu_j)\}_{j=1}^{S}$ amounts to recovering the S-sparse vector $\mathbf{b}$ from the measurement vector $\mathbf{y}$. A—by now standard—recovery approach is to solve a convex ℓ_1-norm-minimization program. From [22, Thm. 5.1] we know that, provided the x_ℓ are i.i.d. sub-Gaussian random variables, and provided that $S \leq cL/(\log L)^4$ for a sufficiently small numerical constant c, with high

probability, all S-sparse vectors $\mathbf{b}$ can be recovered from $\mathbf{y}$ via ℓ_1-minimization. Note that the result [22, Thm. 5.1] only applies to the Gabor matrix $\mathbf{G_x}$ and therefore does not apply to the super-resolution problem where the "columns" $\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}$ are highly correlated. In fact, the two problems are conceptually very different: [22, Thm. 5.1] shows that the columns of the Gabor matrix $\mathbf{G_x}$ are nearly orthogonal, while the "columns" $\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}$ are extremely correlated for two pairs of time-frequency shifts that are close.

7.2.4   Only time or only frequency shifts

Next, suppose we only have time- or only frequency shifts. In both cases, recovery of the unknowns $\{(b_j, \tau_j)\}$ and $\{(b_j, \nu_j)\}$, respectively, is equivalent to the recovery of a weighted superposition of spikes from low-frequency samples. Specifically, if we only have frequency shifts, and therefore $\tau_j = 0$ for all j, the input-output relation (7.3) reduces to

$$y_p = x_p \sum_{j=1}^{S} b_j e^{i2\pi p \nu_j}, \quad p = -N, ..., N. \tag{7.6}$$

Note that the samples $\{y_p\}$ above are samples of a mixture of S complex sinusoids, and estimation of the coefficients $\{(b_j, \nu_j)\}$ corresponds to determining the magnitudes and the frequency components of these sinusoids. Estimating the coefficients $\{(b_j, \nu_j)\}$ from the samples $\{y_p\}$ is known as a line spectral estimation problem and can be solved using classical approaches such as Prony's method [23, Ch. 2], as well as convex programming based approaches [10]. An analogous situation arises when there are only time shifts ($\nu_j = 0$ for all j) as taking the discrete Fourier transform of y_p yields a relation exactly equal to equation (7.6).

7.3   Atomic norm minimization and associated performance guarantees

We next present a convex optimization based recovery algorithm. Even though the corresponding convex program can be solved in polynomial time, standard solvers are currently computationally very expensive, limiting the practical applicability. However, in Section 7.4 we discuss a very closely related convex program that has a significantly better computational efficiency at the cost of making a small griding error that is due to a discretization step. Since the results and intuition for both approaches are nearly the same, we start with discussing the continuous case in this section.

7.3.1   Atomic norm minimization

We first define for convenience the vector $\mathbf{r}_j := [\tau_j, \nu_j]$, and write the input-output relation (7.3) in matrix-vector form:

$$\mathbf{y} = \mathbf{G_x}\mathbf{F}^H\mathbf{z}, \quad \mathbf{z} = \sum_{j=1}^{S} b_j \mathbf{f}(\mathbf{r}_j). \tag{7.7}$$

Here, $\mathbf{F}^H \in \mathbb{C}^{L^2 \times L^2}$ is the (inverse) 2D discrete Fourier transform matrix with the entry in the (k, ℓ)-th row and (r, q)-th column given by $[\mathbf{F}^H]_{(k,\ell),(r,q)} := \frac{1}{L^2} e^{i2\pi \frac{qk+r\ell}{L}}$ and the entries of the vector $\mathbf{f}$ are given by $[\mathbf{f}(\mathbf{r})]_{(r,q)} := e^{-i2\pi(r\tau+q\nu)}$, $k, \ell, q, r = -N, \ldots, N$ (here, and in the following we use for convenience a two or three dimensional index to refer to entries of vectors and matrices). Moreover, $\mathbf{G_x} \in \mathbb{C}^{L \times L^2}$ is the Gabor matrix defined in equation (7.5).

The significance of the representation in (7.7) is that recovery of the unknowns $\{(b_j, \mathbf{r}_j)\}$ from $\mathbf{z}$ is a 2D line spectral estimation problem that can be solved with standard spectral estimation techniques such as Prony's method [9]. Therefore, we only need to recover $\mathbf{z} \in \mathbb{C}^{L^2}$ from $\mathbf{y} \in \mathbb{C}^L$. To do so, we use that $\mathbf{z}$ is a sparse linear combination of atoms in the set $\mathcal{A} := \{\mathbf{f}(\mathbf{r}), \mathbf{r} \in [0,1]^3\}$. A regularizer that promotes such a sparse linear combination is the atomic norm induced by these signals [24], defined as

$$\|\mathbf{z}\|_{\mathcal{A}} := \inf_{b_k \in \mathbb{C}, \mathbf{r}_k \in [0,1]^2} \left\{ \sum_k |b_k| : \mathbf{z} = \sum_k b_k \mathbf{f}(\mathbf{r}_k) \right\}.$$

We estimate $\mathbf{z}$ by solving the basis pursuit type atomic norm minimization problem problem

$$\mathrm{AN}(\mathbf{y}): \quad \underset{\tilde{\mathbf{z}}}{\text{minimize}} \, \|\tilde{\mathbf{z}}\|_{\mathcal{A}} \text{ subject to } \mathbf{y} = \mathbf{A}\tilde{\mathbf{z}}. \tag{7.8}$$

To summarize, we estimate the attenuation factors b_k and time-frequency shifts $\mathbf{r}_k$ from $\mathbf{y}$ by

i. solving $\mathrm{AN}(\mathbf{y})$ in order to obtain $\mathbf{z}$,
ii. estimating the $\mathbf{r}_k$ from $\mathbf{z}$ by solving the corresponding 2D-line spectral estimation problem, and
iii. solving the linear system of equations $\mathbf{y} = \sum_{k=0}^{S-1} b_k \mathbf{A}\mathbf{f}(\mathbf{r}_k)$ for the b_k.

We remark that the $\mathbf{r}_k$ may be obtained more directly from a solution to the dual of (7.8) [14, Sec. 6], see also [12, Sec. 3.1], [10, Sec. 4], [11, Sec. 2.2] for details on this approach applied to related problems.

Since computation of the atomic norm involves taking the infimum over infinitely many parameters, finding a solution to $\mathrm{AN}(\mathbf{y})$ may appear to be daunting. For the one-dimensional case (i.e., only time or frequency shifts), the atomic norm can be characterized in terms of linear matrix inequalities [11, Prop. 2.1]. This characterization is based on the Vandermonde decomposition lemma for

Toeplitz matrices, and allows to formulate the atomic norm minimization program as a semidefinite program that can be solved in polynomial time. While this lemma generalizes to higher dimensions [25, Thm. 1], it fundamentally comes with a rank constraint that appears to prohibit an straightforward characterization of the atomic norm in terms of linear matrix inequalities. Nevertheless, based on [25, Thm. 1], one can obtain a semidefinite programming *relaxation* of $\mathrm{AN}(\mathbf{y})$, which can be solved in polynomial time. Similarly, a solution of the dual of $\mathrm{AN}(\mathbf{y})$ can be found with a semidefinite programming relaxation. Since the computational complexity of the corresponding semidefinite programs is quite large, we will not dive into the details of those semidefinite programming relaxations. As mentioned before, instead, we show in Section 7.4 that the parameters $\{\mathbf{r}_j\}$ can be recovered on an arbitrarily fine grid via ℓ_1-minimization. While this leads to a gridding error, the grid may be chosen sufficiently fine for the gridding error to be negligible compared to the error induced by additive noise, and in practice, there is always some additive noise.

7.3.2 Recovery guarantees for atomic norm minimization

We consider a random probing signal by taking the samples of the probing signal x_ℓ in (7.3) to be i.i.d. Gaussian random variables. More generally, the result presented below continues to hold if we choose the samples as sub-Gaussian random variables, for example as random signs. Note that the probing signal can be chosen by the radar engineer, therefore, choosing the coefficients at random is not problematic and can be done in practice. The theorem stated below shows that, with high probability, the triplets $\{(b_j, \tau_j, \nu_j)\}$ can be recovered perfectly from the samples by solving a convex program, provided that the number of time-frequency shifts is sufficiently smaller than the number of measurement, and provided that the following minimum separation condition holds:

DEFINITION 7.1 (Minimum separation condition) We say the time-frequency shifts $(\tau_j, \nu_j) \in [0, 1]^2, j = 1, \ldots, S$, satisfy the minimum separation condition if

$$\max(|\tau_j - \tau_{j'}|, |\nu_j - \nu_{j'}|) \geq \frac{2.38}{N}, \text{ for all } j \neq j', \tag{7.9}$$

where $|\tau_j - \tau_{j'}|$ is the wrap-around distance on the unit circle. For example, $|3/4 - 1/2| = 1/4$ but $|5/6 - 1/6| = 1/3 \neq 2/3$.

Note that the time-frequency shifts must not be separated in both time *and* frequency, for example the minimum separation condition can hold even when $\tau_j = \tau_{j'}$ for some $j \neq j'$. The main result on recovery via atomic norm minimization from the paper [14] is stated next.

THEOREM 7.2 *Assume that the samples of the probing signal* $\mathbf{x} \in \mathbb{C}^L$ *are i.i.d.* $\mathcal{N}(0, 1/L)$ *random variables. Consider a signal where the sign of the attenuation factors* $\{b_j\}_{j=1}^S$ *is i.i.d. uniform on* $\{-1, 1\}$ *or the complex unit disc, and suppose that the time-frequency shifts* $\{(\tau_j, \nu_j)\}_{j=1}^S$ *obey the minimum separation*

condition. Furthermore, choose $\delta > 0$ and assume that the number of non-zero attenuation factors, S, and the number of measurements, L, obey

$$S \leq c \frac{L}{\log^3(L/\delta)},$$

where c is a numerical constant. Then, with probability at least $1 - \delta$, $\mathbf{z} = \sum_{j=1}^{S} b_j \mathbf{f}(\mathbf{r}_j)$ is the unique minimizer of $\mathrm{AN}(\mathbf{y})$, $\mathbf{y} = \mathbf{G_x}\mathbf{F}^H\mathbf{z}$.

This result is essentially optimal in terms of the allowed sparsity level, as the number S of unknowns can be linear—up to a logarithmic-factor—in the number of observations L. Even when we are given the time-frequency shifts (τ_j, ν_j), we can only hope to recover the corresponding attenuation factors b_j by solving the linear system of equations in (7.3), provided that $S \leq L$.

Since the complex-valued coefficients b_j in the radar model describe the attenuation factors, it is natural to assume that the phases of different b_j are independent from each other and are uniformly distributed on the unit circle of the complex plane. Indeed, standard models for wireless communication channels and radar [26], assume the coefficients $\{b_j\}$ to be complex Gaussian distributed. Nevertheless, we believe that the random sign assumption is not necessary for Theorem 7.2 to hold. In Section 7.7, we discuss a closely related problem which does not require the random sign assumption, and thus provides a basis for the claim of the random sign assumption not being necessary. Finally, we would like to point out that Theorem 7.2 continues to hold for sub-Gaussian sequences x_ℓ, for example random signs.

7.3.3   Necessity of minimum separation

Theorem 7.2 imposes a minimum-separation condition, and indeed some form of separation between the time-frequency shifts is necessary for *stable* recovery. To be specific, we consider the simpler problem of line spectral estimation (see Section 7.2.4) that is obtained from our setup by setting $\tau_j = 0$ for all j. Clearly, any condition necessary for the line spectral estimation problem is also necessary for the super-resolution radar problem.

If there are S' frequencies $\{\nu_k\}_{k=1}^{S'}$ in an interval of length smaller than $\frac{2S'}{L}$, and S' is large, then in the presence of even a tiny amount of noise, stable recovery of the attenuation factors and time-frequency shifts even from

$$\mathbf{z} = \sum_{j=1}^{S} b_j \mathbf{f}(\mathbf{r}_j),$$

where we set, with a slight abuse of notation, $[\mathbf{f}(\nu_j)] = e^{i2\pi j\nu_j}$, is not possible (see [27, Thm. 1.1] and [10, Sec. 1.7]). Condition (7.9) allows us to have $0.4\,S'$ time-frequency shifts in an interval of length $\frac{2S'}{L}$, which is optimal up to the constant 0.4.

This argument illustrates that for stable recovery, it is relevant whether a number of frequencies, say S many, cluster together in a small interval of size smaller than S/L. To illustrate this point further, consider again the simpler problem of line spectral estimation, i.e., recovery or the frequencies from $\mathbf{z} = \sum_{j=1}^{S} b_j \mathbf{f}(\nu_j)$, with $[\mathbf{f}(\nu_j)] = e^{i2\pi j \nu_j}$. Consider the following (Vandermonde) matrix parameterized by S and ϵ:

$$\mathbf{V} = [\mathbf{f}(0), \mathbf{f}((1-\epsilon)/L), \dots, \mathbf{f}_L((2S-1)(1-\epsilon)/L)].$$

We next state a theorem, which lower bounds the condition number of $\mathbf{V}$. The lower bound implies that there are signals with S many frequencies in an interval smaller than S/L, that are indistinguishable even under a tiny amount of additive noise.

THEOREM 7.3 ([28, Thm. 1.3])   *Fix some $\epsilon \in (0,1)$, let $K = \frac{1}{1-\epsilon} L$, and let $S = O(\log(L/(1-\epsilon)))$. Then the matrix $\mathbf{V}$ has condition number at least $e^{O(\epsilon S)}$.*

The theorem implies that there exists a vector $\mathbf{b}$ with unit norm that obeys $\|\mathbf{V}\mathbf{b}\|_2 \le e^{-O(\epsilon S)}$. As a consequence,

$$\sum_{\ell=-N}^{N} \left(\sum_{j \text{ odd}} b_j e^{i2\pi \nu_j \ell} + \sum_{j \text{ even}} b_j e^{i2\pi \nu_j \ell} \right)^2 = \|\mathbf{V}\mathbf{b}\|_2^2 \le e^{-O(\epsilon S)},$$

which means that there are two sets of S many point sources each, with separation $\frac{2(1-\epsilon)}{L}$, but telling them apart requires an exponentially small additive error. To obtain intuition on the constants involved in the statement, we plot the condition number of $\mathbf{V}$ for different values of S and ϵ in Figure 7.1.
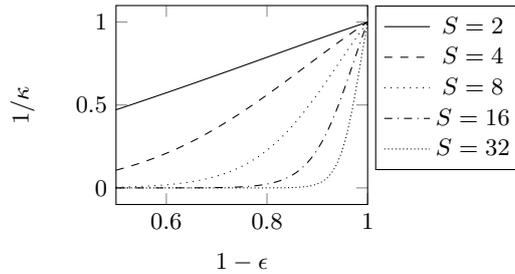

Figure 7.1 Inverse of the condition number κ of the matrix $\mathbf{V}$ with entries $[\mathbf{V}]_{pq} = e^{-i2\pi pq(1-\epsilon)/L}$, $L = 200$, and $q = 1, \dots, S$, for different values of the number of sources S as a function of the separation between frequencies of $(1-\epsilon)/L$.

While those two arguments show that some form of separation between the time-frequency shifts is necessary, the exact form of separation required in (7.9) may not be necessary for stable recovery and less restrictive conditions may suffice. Indeed, in the simpler problem of line spectral estimation, Donoho [27] showed that stable super-resolution is possible via an exhaustive search algorithm

even when condition (7.9) is violated locally, as long as every interval of the ν-axis of length $\frac{2S'}{L}$ contains less than $S'/2$ frequencies and S' is small (in practice, think of $S' \lesssim 10$).

7.3.4 Implications for the detection accuracy of radar systems

Translated to the continuous setup, Theorem 7.2 implies that with high probability, the triplets $(b_j, \bar{\tau}_j, \bar{\nu}_j)$ can be identified perfectly provided that

$$|\bar{\tau}_j - \bar{\tau}_{j'}| \geq \frac{4.77}{B} \quad \text{or} \quad |\bar{\nu}_j - \bar{\nu}_{j'}| \geq \frac{4.77}{T}, \quad \text{for all } j \neq j', \qquad (7.10)$$

and provided that $S \leq c\frac{BT}{\log^3(BT)}$. Since we can *exactly* identify the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$, as opposed to only localizing them on a grid, this result offers a significant improvement in resolution over conventional radar techniques. Specifically, with a standard pulse-Doppler radar which samples the received signal and performs digital matched filtering in order to detect the objects, the delay-Dopper shifts $(\bar{\tau}_j, \bar{\nu}_j)$ can only be estimated up to an uncertainty of about $(1/T, 1/B)$.

We hasten to add that in the radar literature, the term super-resolution is often used for the ability to resolve objects that are very close—even closer than the Rayleigh resolution limit [29] that is proportional to $1/B$ and $1/T$ for delay and Doppler resolution, respectively. The norm minimization approach discussed here permits identification of *each* object with precision much higher than $1/B$ and $1/T$ as long as the other objects are not too close, specifically other objects should be separated by a constant multiple of the Rayleigh resolution limit as formalized by the minimum separation condition (7.10). Recall that, however, any method that attempts to recover objects closer than the resolution limit can only succeed if there are very few objects below that limit, since resolving many objects that are all below the resolution limit is in general impossible, as discussed previously in Section 7.3.3.

Finally, recall that the approach discussed here allows the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ to lie in $[-T/2, T/2] \times [-B/2, B/2]$ so the delay-Doppler pairs can lie in a rectangle of area $L = BT \gg 1$. The ability to handle a potentially large region in which delay-Doppler pairs can lie in is important in radar applications, since we might need to resolve objects with large relative distances and relative velocities.

7.3.5 Can standard non-parametric methods yield similar performance?

We finally note that standard non-parametric estimation methods such as the MUSIC algorithm can in general *not* be applied directly to the super-resolution radar problem. The reason is that MUSIC relies on *multiple* measurements (often referred to as snapshots) [9, Sec. 6.3], whereas we assume only a *single* measurement $\{y_p\}_{p=-N}^{N}$ to be available. In our context, multiple measurements would

amount to carrying out multiple, independent input-output measurements. However, by choosing the probing signal $\mathbf{x}$ in (7.3) to be periodic, a single measurement can be transformed into multiple measurements and for that case, for example the MUSIC algorithm may be applied. However, this approach, discussed in detail in [14, Appendix H], requires the frequencies $\{\nu_j\}$ to be distinct, the time-shifts $\{\tau_j\}$ to lie in a significantly smaller range than $[0,1]$, and $S < \sqrt{L}$, as opposed to the much milder condition $S < L/\log^3(L/\delta)$ required by the convex program. In addition, applying MUSIC in that way is (significantly) more sensitive to noise than the convex programming based approach discussed in this chapter. If multiple measurements are available, for example by observing distinct paths of a signal by an array of antennas, the situation might be different. For that case, subspace methods have been studied for delay-Doppler estimation in the paper [30].

7.4   Super-resolution radar on a fine grid

An practical approach to estimate the triplets $\{(b_j, \tau_j, \nu_j)\}$ from the received signal $\mathbf{y}$ in the input-output relation (7.3) is to suppose the time-frequency shifts lie on a *fine* grid, and solve the problem on that grid. In general this leads to a gridding error, which, however, becomes small as the grid gets finer [31]. We next discuss the corresponding (discrete) sparse signal recovery problem.

Suppose the time-frequency shifts lie on a *fine* grid with spacing $(1/K, 1/K)$, where K is an integer obeying $K \geq L$. Let $\mathbf{b} \in \mathbb{C}^{K^2}$ be the signal with each entry $b_{m,n}$ corresponding to one of the grid points, with non-zeros equal to the attenuation factors b_j for the time-frequency shifts $(\tau_j, \nu_j), j = 1, \ldots, S$. See Figure 7.2 for an illustration. With this assumption, the input-output relation (7.3) becomes:

$$y_p = \sum_{m,n=0}^{K-1} b_{m,n}[\mathcal{F}_{m/K}\mathcal{T}_{n/K}\mathbf{x}]_p, \quad p = -N, \ldots, N.$$

Writing this relation in matrix-vector form yields

$$\mathbf{y} = \mathbf{R}\mathbf{b},$$

where, as before, $\mathbf{y}$ is the vector containing as entries the values y_p, and $\mathbf{R} \in \mathbb{C}^{L \times K^2}$, is the matrix with (m, n)-th column given by $\mathcal{F}_{m/K}\mathcal{T}_{n/K}\mathbf{x}$. The matrix $\mathbf{R}$ contains as columns "fractional" time-frequency shifts of the sequence x_ℓ. If $K = L$, $\mathbf{R}$ contains as columns only "whole" time-frequency shifts of $\mathbf{x}$ and $\mathbf{R}$ is equal to the Gabor matrix $\mathbf{G_x}$ defined by (7.5). In this sense, $K = L$ is the natural grid (see Section 7.2.3) and the ratio $\mathrm{SRF} := K/L$ can be interpreted as a super-resolution factor. The super-resolution factor determines by how much the $(1/K, 1/K)$ grid is finer than the original $(1/L, 1/L)$ grid.

A standard approach to the recovery of the sparse signal $\mathbf{b}$ from the underdetermined linear system of equations $\mathbf{y} = \mathbf{R}\mathbf{b}$ is to solve the following convex
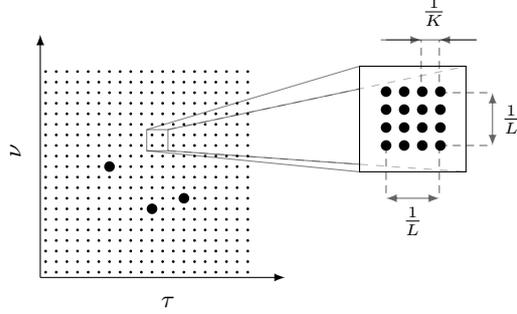
Figure 7.2 Time frequency shifts that lie on a grid: $(1/L, 1/L)$ is the "natural" grid, and $(1/K, 1/K)$ is the fine grid. Each dot corresponds to a potential non-zero, and the larger dots correspond to the actual non-zeros $\{b_j\}$.

program:

$$\text{L1}(\mathbf{y})\colon \quad \underset{\tilde{\mathbf{b}}}{\text{minimize}} \ \left\| \tilde{\mathbf{b}} \right\|_1 \text{ subject to } \mathbf{y} = \mathbf{R}\tilde{\mathbf{b}}. \tag{7.11}$$

The following theorem is the main result from [14] for recovery on the fine grid.

THEOREM 7.4   *Assume that the samples of the probing signal* $\mathbf{x}$ *are i.i.d.* $\mathcal{N}(0, 1/L)$ *random variables,* $L = 2N + 1$, *and that* $L = 2N + 1 \geq 1024$. *Consider a signal* $\mathbf{b}$ *supported on* $\mathcal{S} \subseteq \{0, ..., K-1\}^2$, *and suppose that it satisfies the minimum separation condition*

$$\min_{(m,n),(m',n')\in\mathcal{S}\colon (m,n)\neq(m',n')} \frac{1}{K} \max(|m - m'|, |n - n'|) \geq \frac{2.38}{N}.$$

Moreover, suppose that the non-zeros of $\mathbf{b}$ *are i.i.d. uniform on* $\{-1, 1\}$ *or the complex unit disk. Choose* $\delta > 0$, *let* $\mathbf{y} = \mathbf{R}\mathbf{b}$ *be the measurement corresponding to* $\mathbf{b}$, *and suppose that the number of non-zeros of* $\mathbf{b}$ *is sufficiently smaller than the number of samples* L

$$S \leq c\frac{L}{\log^3(L/\delta)},$$

where c *is a numerical constant. Then, with probability at least* $1 - \delta$, $\mathbf{b}$ *is the unique minimizer of* $\text{L1}(\mathbf{y})$, $\mathbf{y} = \mathbf{R}\mathbf{b}$.

Note that Theorem 7.4 does not impose any restriction on K, in particular it can be arbitrarily large. The proof of Theorem 7.4, discussed in Section 7.5.2 is closely linked to that of Theorem 7.2.

7.4.1   Implementation details

The matrix $\mathbf{R}$ has dimension $L \times K^2$, thus as the grid becomes finer, i.e., K becomes larger, the complexity of solving $\text{L1}(\mathbf{y})$ increases. The complexity of

solving L1($\mathbf{y}$) can be managed as follows. First, the complexity of first-order convex optimization algorithms (such as TFOCS [32]) for solving L1($\mathbf{y}$) is dominated by multiplications with the matrices $\mathbf{R}$ and $\mathbf{R}^H$. Due to the structure of $\mathbf{R}$, those multiplications can be done very efficiently utilizing the fast Fourier transform. Second, in practice we have $(\bar{\tau}_j, \bar{\nu}_j) \in [0, \tau_{\max}] \times [0, \nu_{\max}]$, which means that

$$(\tau_j, \nu_j) \in \left[0, \frac{\tau_{\max}}{T}\right] \times \left[0, \frac{\nu_{\max}}{B}\right]. \tag{7.12}$$

It is therefore sufficient to consider the restriction of $\mathbf{R}$ to the $\frac{\tau_{\max}\nu_{\max}K^2}{BT} = \tau_{\max}\nu_{\max}L \cdot \mathrm{SRF}^2$ many columns corresponding to the time-frequency shifts (τ_j, ν_j) satisfying (7.12). Since typically $\tau_{\max}\nu_{\max} \ll BT = L$, this results in a significant reduction of the problem size.

7.4.2   Numerical results and robustness

We next discuss numerical results which show that the convex optimization based super-resolution approach is robust to noise. Consider the following modification of ℓ_1-norm minimization, which accounts for noise and the gridding error:

$$\text{L1-ERR}: \underset{\tilde{\mathbf{b}}}{\text{minimize}} \left\| \tilde{\mathbf{b}} \right\|_1 \text{ subject to } \left\| \mathbf{y} - \mathbf{R}\tilde{\mathbf{b}} \right\|_2^2 \le \delta.$$

The parameter δ is chosen on the order of the noise variance or the expected gridding error.

The paper [14] considered a synthetic problem with $L = 201$, where each problem instance is generated by drawing $S = 10$ time-frequency shifts (τ_j, ν_j) uniformly at random from the interval $[0, 2/\sqrt{201}]^2$. This amounts to drawing the corresponding delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ from the interval $[0, 2] \times [0, 2]$. The attenuation factors b_j corresponding to the time-frequency shifts were drawn uniformly at random from the complex unit disc, independently across j. Measurements were then obtained according to the input-output relation (7.3).

Figure 7.6 depicts the average resolution error versus the super-resolution factor $\mathrm{SRF} = K/L$. The resolution error is defined as the average over $j = 1, \ldots, S$ of $L\sqrt{(\hat{\tau}_j - \tau_j)^2 + (\hat{\nu}_j - \nu_j)^2}$, where the $(\hat{\tau}_j, \hat{\nu}_j)$ are the estimates of the time-frequency shifts extracted from a solution of L1-ERR, obtained with the SPGL1 solver [33]. Note that the resolution attained at $\mathrm{SRF} = 1$ corresponds to the resolution attained by matched filtering and by the compressive sensing radar architecture [4] discussed in Section 7.2.3.

As mentioned before, there are two error sources incurred by this approach. The first is the gridding error obtained by assuming the points lie on a fine grid with grid constant $(1/K, 1/K)$, which decays in K. The second is the additive noise error, which is constant. The figure shows that for SRF larger than one, the resolution is significantly improved using the super-resolution radar approach. Moreover we see that for small super-resolution factors SRF, the gridding error dominates, while for large values of SRF, the additive noise error dominates. In
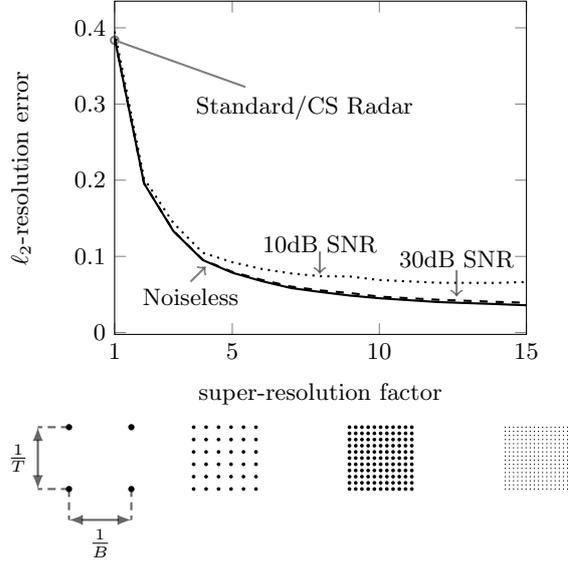
Figure 7.3 Super-resolution radar uniformly provides better resolution error than standard or CS Radar. The plot show the resolution error for the recovery of $S = 10$ time-frequency shifts from the observation $\mathbf{y}$ with and without additive Gaussian noise $\mathbf{n}$ of a certain signal-to-noise ratio $\text{SNR} := \|\mathbf{y}\|_2 / \|\mathbf{n}\|_2$. The resolution error is defined as the average over $L\sqrt{(\hat{\tau}_j - \tau_j)^2 + (\hat{\nu}_j - \nu_j)^2}$, where (τ_j, ν_j) are the original time-frequency shifts, and the $(\hat{\tau}_j, \hat{\nu}_j)$ are the time-frequency shifts on the grid, obtained by solving L1-ERR, for different super-resolution factors. The different super-resolution factors are illustrated below the plot.

this experiment, the gridding error approximately decays as 1/SRF. The experiment demonstrates that in practice solving the super-resolution radar problem on a fine grid is essentially as good as solving it on the continuum—provided the super-resolution factor is chosen sufficiently large, so that the gridding error becomes negligible relative to the error due to additive noise.

7.5   Proof outline

In this section, we discuss the proofs of Theorems 7.2 and 7.4 from [14] which are closely linked. Specifically, both Theorems 7.2 and 7.4 follow from the existence of certain dual certificates, and the certificate for proving Theorem 7.4 is obtained directly from the certificate constructed to prove Theorem 7.2.

7.5.1   Proof of Theorem 7.2

Theorem 7.2 is proven by constructing an appropriate dual certificate; the existence of this certificate guarantees that the solution to $\text{AN}(\mathbf{y})$ is $\mathbf{z}$, as formalized by Proposition 1 below. Proposition 1 is a consequence of strong duality, and is

well known for the discrete setting from the compressed sensing literature [3]. The proof is standard, see for example [11, Proof of Prop. 2.4]. For convenience, in this section, we set $\mathbf{A} = \mathbf{G_x}\mathbf{F}^H$, so that $\mathbf{y} = \mathbf{Az}$.

Proposition 1   Let $\mathbf{y} = \mathbf{Az}$ with $\mathbf{z} = \sum_{j=1}^{S} b_j\mathbf{f}(\mathbf{r}_j)$. Suppose there exists a function, called dual certificate, of the form $Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{Af}(\mathbf{r}) \rangle$, parameterized by the complex coefficients $\mathbf{q} \in \mathbb{C}^L$, such that

$$Q(\mathbf{r}_j) = \mathrm{sign}(b_j), \text{ for all } j, \text{ and}$$
$$|Q(\mathbf{r})| < 1 \text{ for all } \mathbf{r} \in [0,1]^2 \setminus \{\mathbf{r}_1, \dots, \mathbf{r}_S\}. \tag{7.13}$$

Moreover, suppose that the vectors $\{\mathbf{Af}(\mathbf{r}_j)\}_{j=1}^{S}$, are linearly independent. Then $\mathbf{z}$ is the unique minimizer of AN($\mathbf{y}$).

A condensed argument showing that Proposition 1 is true is the following. Suppose for contradiction that there exists another optimal solution $\mathbf{z}' = \sum_j b'_j\mathbf{f}(\mathbf{r}'_j)$ with $\|\mathbf{z}'\|_{\mathcal{A}} = \sum_j |b'_j|$ and $\{\mathbf{r}'_j\} \neq \{\mathbf{r}_j\}$. First suppose that $\{\mathbf{r}'_j\} \subseteq \{\mathbf{r}_j\}$. Then, linear independence of the vectors $\{\mathbf{Af}(\mathbf{r}_j)\}$ contradicts that $\mathbf{z}' \neq \mathbf{z}$. Next, suppose that $\{\mathbf{r}'_j\} \nsubseteq \{\mathbf{r}_j\}$. We then have that

$$\|\mathbf{z}'\|_{\mathcal{A}} - \|\mathbf{z}\|_{\mathcal{A}} = \sum_j |b'_j| - \sum_j |b_j| \overset{i}{>} \mathrm{Re}\sum_j Q^*(\mathbf{r}'_j)b'_j - \mathrm{Re}\sum_j Q^*(\mathbf{r}_j)b_j$$
$$\overset{ii}{=} \mathrm{Re}\sum_j \mathbf{q}^H\mathbf{Af}(\mathbf{r}'_j)b'_j - \mathrm{Re}\sum_j \mathbf{q}^H\mathbf{Af}(\mathbf{r}_j)b_j$$
$$\overset{iii}{=} 0.$$

Here, inequality *i*) follows from the dual polynomial interpolating the sign pattern, and from $|Q(\mathbf{r}'_j)| < 1$ for at least one $\mathbf{r}'_j$, which in turn follows from $\{\mathbf{r}'_j\} \nsubseteq \{\mathbf{r}_j\}$ and $|Q(\mathbf{r})| < 1$ for all $\mathbf{r} \notin \{\mathbf{r}_j\}$, by assumption. Inequality *ii*) follows from the definition of the dual polynomial, and inequality *iii*) follows from $\mathbf{Az}' = \mathbf{Az}$, by assumption. This contradicts that $\mathbf{z}'$ is an optimal solution.

We now turn to the construction of a dual certificate obeying the conditions of Proposition 1, which concludes the proof of Theorem 7.2. The construction of the dual certificate Q is inspired by the construction of related certificates in [10, 11]. First, recall that the entries of $\mathbf{f}(\mathbf{r}), \mathbf{r} = [\tau, \nu]$, are given by $[\mathbf{f}(\mathbf{r})]_{(k,p)} = e^{i2\pi(k\tau + p\nu)}$. From

$$Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{Af}(\mathbf{r}) \rangle = \langle \mathbf{A}^H\mathbf{q}, \mathbf{f}(\mathbf{r}) \rangle,$$

it is seen that Q is a two-dimensional trigonometric polynomial in the variables τ and ν with coefficient vector $\mathbf{A}^H\mathbf{q}$. To build the certificate, we therefore need to construct a two-dimensional trigonometric polynomial that satisfies the interpolation and boundedness condition (7.13), and whose coefficients are constraint to be of the form $\mathbf{A}^H\mathbf{q}$. Since the matrix $\mathbf{A}$ is a function of the random probing signal $\mathbf{x}$, Q is a *random* trigonometric polynomial. We construct Q explicitly.

It is instructive to first consider the construction of a *deterministic* two-

dimensional trigonometric polynomial

$$\bar{Q}(\mathbf{r}) = \langle \bar{\mathbf{q}}, \mathbf{f}(\mathbf{r}) \rangle$$

with unconstraint, deterministic coefficients $\bar{\mathbf{q}} \in \mathbb{C}^L$, that satisfies the interpolation and boundedness conditions (7.13), but whose coefficient vector $\bar{\mathbf{q}}$ is *not* constraint to be of the form $\mathbf{A}^H \mathbf{q}$. Such a construction has been established—provided the the parameters $\{\mathbf{r}_k\}$ obey the minimum separation condition in Definition 7.9—by Candès and Fernandez-Granda [10, Prop. 2.1, Prop. C.1] for the one- and two-dimensional case. To construct the polynomial $\bar{Q}$, [10] interpolate the signs $\{\text{sign}(b_j)\}$ with a fast-decaying kernel

$$\bar{G}(\mathbf{r}) := F(\tau) F(\nu),$$

and slightly adopt this interpolation near the parameters $\{\mathbf{r}_j\}$ with the partial derivatives $\bar{G}^{(n_1, n_2)}(\mathbf{r}) := \frac{\partial^{n_1}}{\partial \tau^{n_1}} \frac{\partial^{n_2}}{\partial \nu^{n_2}} \bar{G}(\mathbf{r})$ to ensure that local maxima are achieved at the $\mathbf{r}_j$:

$$\bar{Q}(\mathbf{r}) = \sum_{j=1}^{S} \bar{\alpha}_j \bar{G}(\mathbf{r} - \mathbf{r}_j) + \bar{\alpha}_{1j} \bar{G}^{(1,0)}(\mathbf{r} - \mathbf{r}_j) + \bar{\alpha}_{2j} \bar{G}^{(0,1)}(\mathbf{r} - \mathbf{r}_j). \qquad (7.14)$$

Here, F is the squared Fejér kernel which is a particular trigonometric polynomial with coefficients g_j, i.e.,

$$F(t) = \sum_{j=-N}^{N} g_j e^{i2\pi t j}.$$

Shifted versions of the polynomial F (i.e., $F(t - t_0)$ for some $t_0 \in \mathbb{R}$) and the derivatives of the polynomial F are also one-dimensional trigonometric polynomials of degree N. Therefore $\bar{G}$, its partial derivatives, and shifted versions thereof are two-dimensional trigonometric polynomials of the form $\langle \bar{\mathbf{q}}, \mathbf{f}(\mathbf{r}) \rangle$. The construction of $\bar{Q}$ is concluded by showing that the coefficients $\bar{\alpha}_j, \bar{\alpha}_{1j}, \bar{\alpha}_{2j}, \bar{\alpha}_{3j}$, can be chosen such that $\bar{Q}$ reaches global maxima at the parameters $\{\mathbf{r}_j\}$.

The construction of Q in the paper [14] for proving Theorems 1 follows a similar program. Specifically, the polynomial Q is constructed such that it interpolates the signs $\text{sign}(b_j)$ at $\mathbf{r}_j$ with the functions $G_{\mathbf{n}}(\mathbf{r}, \mathbf{r}_j) = \langle \mathbf{A}\mathbf{g}_{\mathbf{n}}(\mathbf{r}_j), \mathbf{A}\mathbf{f}(\mathbf{r}) \rangle$. Here, $\mathbf{g}_{\mathbf{n}}(\mathbf{r}), \mathbf{n} = (n_1, n_2)$ is the vector with (v, k)-th coefficient given by

$$g_k g_p (i2\pi k)^{n_1} (i2\pi p)^{n_2} e^{-i2\pi(\tau k + \nu p)},$$

where the $\{g_j\}$ are the coefficients of the squared Fejér kernel F defined above. With this definition, we have

$$\mathbb{E}\left[G_{\mathbf{n}}(\mathbf{r}, \mathbf{r}_j)\right] = \bar{G}^{\mathbf{n}}(\mathbf{r} - \mathbf{r}_j).$$

This follows from

$$\mathbb{E}\left[\mathbf{A}^H \mathbf{A}\right] = \mathbf{F}^H \mathbb{E}\left[\mathbf{G}_{\mathbf{x}}^H \mathbf{G}_{\mathbf{x}}\right] \mathbf{F} = \mathbf{I},$$

where we used $\mathbb{E}\left[\mathbf{G}_{\mathbf{x}}^H \mathbf{G}_{\mathbf{x}}\right] = \mathbf{I}$, shown at the end of this section. Moreover, $G_{\mathbf{n}}(\mathbf{r}, \mathbf{r}_k)$ concentrates around $\bar{G}^{\mathbf{n}}(\mathbf{r} - \mathbf{r}_k)$.

Now, Q is constructed by interpolating the signs $\text{sign}(b_j)$ at $\mathbf{r}_j$ with $G_{(0,0)}(\mathbf{r}, \mathbf{r}_j), j = 1, \ldots, S$, and slightly adopting this interpolation near the points $\{\mathbf{r}_j\}$ with linear combinations of the functions $G_{(1,0)}(\mathbf{r}, \mathbf{r}_j)$ and $G_{(0,1)}(\mathbf{r}, \mathbf{r}_j)$, in order to ensure that local maxima of Q are achieved exactly at the $\mathbf{r}_j$. Specifically, we set

$$Q(\mathbf{r}) = \sum_{j=1}^{S} \alpha_j G_{(0,0)}(\mathbf{r}, \mathbf{r}_j) + \alpha_{1j} G_{(1,0)}(\mathbf{r}, \mathbf{r}_j) + \alpha_{2j} G_{(0,1)}(\mathbf{r}, \mathbf{r}_j). \qquad (7.15)$$

Note that $Q(\mathbf{r})$ is a linear combination of the functions $G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_j)$, and by definition of $G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_j)$ it obeys $\langle \mathbf{A}^H \mathbf{q}, \mathbf{f}(\mathbf{r}) \rangle$, for some $\mathbf{q}$, as desired. The proof is concluded by showing that, with high probability, there exists a choice of coefficients α_j, α_{1j}, and α_{2j} such that Q reaches global maxima at the $\mathbf{r}_j$ and $Q(\mathbf{r}_j) = u_j$, for all j. For this argument to work, the particular choice of $G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_j)$ is crucial; the main ingredients for the argument to work are that $G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_j)$ concentrates around $\bar{G}(\mathbf{r} - \mathbf{r}_j)$, and certain properties of the deterministic functions $\bar{G}$ and $\bar{Q}$.

Proof of $\mathbb{E}\left[\mathbf{G_x}^H \mathbf{G_x}\right] = \mathbf{I}$:
By definition of the Gabor matrix in (7.5), the entry in the (k, ℓ)-th row and (k', ℓ')-th column of $\mathbf{G_x}^H \mathbf{G_x}$ is given by

$$[\mathbf{G_x}^H \mathbf{G_x}]_{(k,\ell),(k',\ell')} = \sum_{p=-N}^{N} x_{p-\ell}^* x_{p-\ell'} e^{-i2\pi \frac{kp}{L}} e^{i2\pi \frac{k'p}{L}}.$$

Noting that $\mathbb{E}[x_\ell] = 0$, we conclude that $\mathbb{E}\left[[\mathbf{G_x}^H \mathbf{G_x}]_{(k,\ell),(k',\ell')}\right] = 0$ for $\ell \neq \ell'$. For $\ell = \ell'$, using the fact that $\mathbb{E}\left[x_{p-\ell}^* x_{p-\ell}\right] = 1/L$, we arrive at

$$\mathbb{E}\left[[\mathbf{G_x}^H \mathbf{G_x}]_{(k,\ell),(k',\ell')}\right] = \frac{1}{L} \sum_{p=-N}^{N} e^{i2\pi \frac{(k'-k)p}{L}}.$$

The latter is equal to 1 for $k = k'$ and 0 otherwise. This concludes the proof of $\mathbb{E}\left[\mathbf{G_x}^H \mathbf{G_x}\right] = \mathbf{I}$.

7.5.2 Proof of Theorem 7.4

The following proposition, which is standard in the compressed sensing literature (see e.g., [3]) states that the existence of a dual polynomial guarantees that L1($\mathbf{y}$) recovers $\mathbf{b}$ from the measurement $\mathbf{y} = \mathbf{Gb}$. The proposition is the discrete analogue of Proposition 1 above.

Proposition 2 Let $\mathbf{y} = \mathbf{Rb}$ and let $\mathcal{S}$ denote the support (i.e., the set of non-zero elements) of $\mathbf{b}$. Assume that the columns of $\mathbf{R}$ corresponding to $\mathcal{S}$ are linearly independent. If there exists a vector $\mathbf{v}$ in the row space of $\mathbf{R}$ with

$$\mathbf{v}_\mathcal{S} = \text{sign}(\mathbf{b}_\mathcal{S}) \quad \text{and} \quad \|\mathbf{v}_{\mathcal{S}^c}\|_\infty < 1, \qquad (7.16)$$

then $\mathbf{b}$ is the unique minimizer of L1($\mathbf{y}$). Here, $\mathbf{v}_\mathcal{S}$ its the vector consisting of the

entries of $\mathbf{v}$ indexed by $\mathcal{S}$, and likewise $\mathbf{v}_{\mathcal{S}^c}$ consists of the entries not indexed by $\mathcal{S}$, i.e., the entries indexed by the complement of $\mathcal{S}$.

Theorem 7.4 now follows directly from the existence of the polynomial Q constructed in the previous section. To see this, define $\mathbf{v}$ as $[\mathbf{v}]_{(m,n)} = Q([m/K, n/K])$ and note that $\mathbf{v}$ satisfies (7.16) since $Q([m/K, n/K]) = \text{sign}(\mathbf{b}_{(m,n)})$ for $(m,n) \in \mathcal{S}$ and $|Q([m/K, n/K])| < 1$ for $(m,n) \notin \mathcal{S}$.

7.6   MIMO radar

In this section, we discuss super-resolution imaging in the context of MIMO radar. A MIMO radar uses multiple transmit antennas to send—typically orthogonal—probing signals simultaneously, and records the reflections from the objects with multiple receive antennas. As shown in this section, a MIMO radar can thereby, in principle, resolve the relative angles in addition to the relative distances and velocities of objects with a single measurement.

To illustrate the principle of a MIMO radar, first consider a radar system with a single transmit and multiple receive antennas, and consider a object that lies in the far field of the radar, so that the reflections of the objects that arrive at the receiver are essentially parallel, as illustrated in Figure 7.4. The reflection from the object must travel an additional distance of $d_R \sin(\theta)$ between the signals received at two adjacent receive antennas. Thus, from an estimate of the angle of arrival we can determine the relative position of a object. The angular resolution that can be achieved as well as the number of objects that can be distinguished increases linearly in the number of receive antennas.
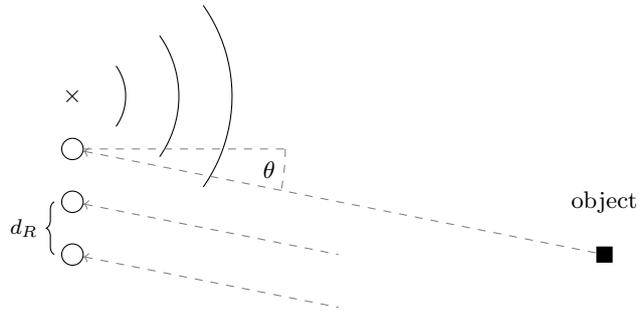

Figure 7.4 Principle of MISO radar: $\times$ and $\bigcirc$ correspond to transmit and receive antennas. The transmit antenna sends a probing signal, and the reflections of the object are received by three receive antennas. Estimating the relative delays of the probing signal allows to estimate the angle of the object relative to the antenna array, in addition to distance and velocity.

As a consequence, for doubling the angular resolution or doubling the number of objects to be distinguishable, a MISO radar need to double its number

of transmit antennas. As we discuss next, using multiple transmit antennas in addition to multiple receive antennas can give a much larger angular resolution from far fewer antennas. Specifically, by arranging N_T transmit and N_R receive antennas in a particular way (see Figure 7.5) a MIMO radar can obtain the same resolution obtained by a MISO (or SIMO) radar with $N_T N_R$ uniformly spaced receive (or transmit) antennas. This is often called a MIMO *virtual array*. In this section, we discuss a MIMO radar model, and show that the fundamental limit for resolving the angle-delay-Doppler triplets is $(1/(N_T N_R), 1/B, 1/T)$. We furthermore show that this limit can be overcome in the sense that triplets can be resolved on a much finer grid, provided they are sufficiently separated.

7.6.1   MIMO Signal model and problem statement

We consider a MIMO radar with N_T transmit and N_R receive antennas that are co-located and lie in a plane along with S objects, see Figure 7.5 for an illustration. The technical results presented in this section generalize to the more general setup where objects lie in three-dimensional space and the transmit and receive antennas lie in a two-dimensional plane. We consider the simpler two-dimensional setup since the generalization to three dimensions are straightforward. As in the previous section, we assume that the objects are located in the far field of the array. As a consequence, propagating waves appear planar and the angles between the object and each antenna are (approximately) the same. We let the transmit and receive antennas be uniformly spaced with spacings $d_T = \frac{1}{2f_c}$ and $d_R = \frac{N_T}{2f_c}$, respectively, where f_c is the carrier frequency of the probing signals. This spacing yields a uniformly spaced *virtual array* with $N_T N_R$ antennas, and thus maximizes the number of virtual antennas achievable with N_T transmit and N_R receive antennas [34, 35]. As explained in Section 7.6.2 below, the (baseband) signal $y_r(t)$ at continuous time t received by antenna $r = 0, \ldots, N_R - 1$, consists of the superposition of the reflections from the objects of the transmitted probing signals $x_j(t), j = 0, \ldots, N_T - 1$, and is given by

$$y_r(t) = \sum_{k=1}^{S} b_k e^{i2\pi r N_T \beta_k} \sum_{j=0}^{N_T - 1} e^{i2\pi j \beta_k} x_j(t - \bar{\tau}_k) e^{i2\pi \bar{\nu}_k t}. \qquad (7.17)$$

Here, $b_k \in \mathbb{C}$ is the attenuation factor, $\beta_k \in [0, 1]$ the angle or azimuth parameter, and $\bar{\tau}_k$ and $\bar{\nu}_k$ are the delay and Doppler shift, all associated with the k-th object. The parameters $\beta_k, \bar{\tau}_k, \bar{\nu}_k$ determine the angle ($\beta = -\sin(\theta)/2$ see Figure 7.5), distance, and velocity of the k-th object relative to the radar. Locating the object therefore amounts to estimating the continuous parameters $b_k, \beta_k, \bar{\tau}_k, \bar{\nu}_k$ from the responses $y_r, r = 0, \ldots, N_R - 1$, to known and suitably selected probing signals x_j.

As discussed in Section 7.2, due to practical constraints, the probing signals must be band-limited and approximately time-limited, and the responses to the probing signals can only be observed over a finite time interval. We make the
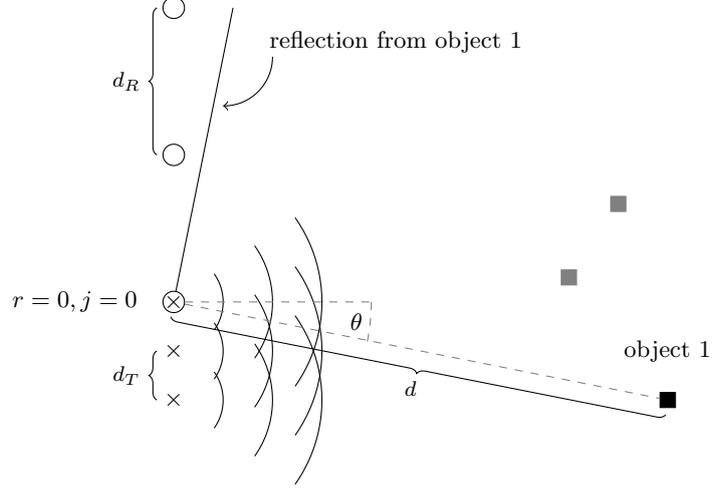
Figure 7.5 Principle of MIMO radar: $\times$ and $\bigcirc$ correspond to transmit and receive antennas. Throughout, we assume the spacing of the N_T transmit and N_R receive antennas to be $d_T = \frac{c}{2f_c}$ and $d_R = \frac{cN_T}{2f_c}$, where f_c is the carrier frequency.

same assumption on the probing signals as well as on the received signals as we made in Section 7.2; in particular we assume that the received signals y_r are observed over an interval of length T and that the probing signals x_j have bandwidth B and are approximately supported on a time interval proportional to T. As explained in Section 7.2, it follows from the band-limitation of the probing signals x_t and the limited observation time of the received signals y_r that the received signals are characterized by the samples

$$[\mathbf{y}_r]_p = \sum_{k=1}^{S} b_k e^{i2\pi r N_T \beta_k} \sum_{j=0}^{N_T - 1} e^{i2\pi j \beta_k} [\mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}_j]_p. \tag{7.18}$$

Here, $\mathbf{y}_r$ contains the samples of the received signal y_r taken at rate $1/B$ (i.e., $[\mathbf{y}_r]_p = y_r(p/B)$ in the interval $p/B \in [-T/2, T/2]$), and $\mathbf{x}_j$ is the vector containing the samples of the probing signal ($[\mathbf{x}_j]_p := x_j(p/B)$).

We have reduced the problem of identifying the locations of the objects under the constraints that the probing signals x_j are band-limited and the responses y_r are observed over a finite time interval only, to the estimation of the parameters $b_k \in \mathbb{C}, (\beta_k, \tau_k, \nu_k) \in [0,1]^3, k = 1, \ldots, S$, from the samples $[\mathbf{y}_r]_p, r = 0, \ldots, N_R - 1, p = -N, \ldots, N$, in the input-output relation (7.18). We call this the super-resolution MIMO radar problem.

7.6.2   Derivation of the MIMO input-output relation

In this section, we derive the MIMO input-output relation (7.17). Towards this goal, we first consider a single object. The j-th antenna transmits the signal

$x_j(t)e^{i2\pi f_c t}$, where f_c is the carrier frequency. This signal propagates to the object, which we assume to be a point scatterer, gets reflected, and propagates back to the r-th receiver. From Figure 7.5, we see that the corresponding delay is, as a function of the angle between antennas and the object, θ, distance to the object, d, and speed of light, c, given by

$$\tilde{\tau} := \frac{2d}{c} + \frac{\sin(\theta)(d_T j + d_R r)}{c} = \bar{\tau} - \beta \frac{2(d_T j + d_R r)}{c}.$$

For the second equality, we defined the angle parameter $\beta := -\sin(\theta)/2$ and the delay $\bar{\tau} := \frac{2d}{c}$. Taking the Doppler shift into account, the reflection of the j-th probing signal received by the r-th receive antenna is given by

$$\tilde{b}x_j(t - \tilde{\tau})e^{i2\pi(f_c + \bar{\nu})(t - \tilde{\tau})}. \tag{7.19}$$

Here, $\tilde{b} \in \mathbb{C}$ is the attenuation factor associated with the object, and $\bar{\nu} := \frac{2v}{c}f_c$ is the Doppler shift, which is a function of the relative velocity, v, of the object. By choosing the antenna spacing as $d_T = \frac{c}{2f_c}$ and $d_R = \frac{cN_T}{2f_c}$, the reflection of the j-th probing signal received by the r-th receive antenna in (7.19) becomes

$$\tilde{b}x_j(t - \bar{\tau})e^{i2\pi(f_c + \bar{\nu})(t - \bar{\tau})}e^{i2\pi(f_c + \bar{\nu})\beta\frac{j + rN_T}{f_c}} \approx \tilde{b}x_j(t - \bar{\tau})e^{i2\pi(f_c + \bar{\nu})(t - \bar{\tau})}e^{i2\pi\beta(j + rN_T)}.$$

Here, the approximation follows by the Doppler shift $\bar{\nu}$ being much smaller than the carrier frequency f_c, therefore $\frac{f_c + \bar{\nu}}{f_c} \approx 1$, and $\tilde{\tau} \approx \bar{\tau}$. If follows that the reflection of the j-th probing signal received by the r-th receive antenna, after demodulation, is

$$bx_j(t - \bar{\tau})e^{i2\pi\bar{\nu}t}e^{i2\pi\beta(j + rN_T)}$$

where we defined $b = \tilde{b}e^{-i2\pi\bar{\nu}\bar{\tau}}$. Next, consider S objects with parameters $(b_k, \beta_k, \bar{\tau}_k, \bar{\nu}_k)$. Since, for S objects, the (demodulated) signal y_r received by antenna r consists of the superposition of the reflections of the probing signals $x_j, j = 0, \ldots, N_T - 1$, transmitted by the transmit antennas, we obtain the input-output relation in equation (7.17) simply by summing over the reflections given by $b_k x_j(t - \bar{\tau}_k)e^{i2\pi\bar{\nu}_k t}e^{i2\pi\beta_k(j + rN_T)}$.

7.6.3   MIMO atomic norm minimization

Recall that our goal is to recover the unknown parameters $(b_k, \beta_k, \tau_k, \nu_k)$ from the measurements $\{\mathbf{y}_r\}$. Towards this goal, we proceed analogously as in Section 7.2.1. We start by defining for convenience the vector $\mathbf{r} := [\beta, \tau, \nu]$, and write the input-output relation (7.18) in matrix-vector form:

$$\mathbf{y} = \mathbf{A}\mathbf{z}, \quad \mathbf{z} = \sum_{k=1}^{S} b_k \mathbf{f}(\mathbf{r}_k). \tag{7.20}$$

Here $\mathbf{y}$ is obtained by stacking the vectors $\mathbf{y}_0^T, \ldots, \mathbf{y}_{N_R-1}^T$, and the vector $\mathbf{f}(\mathbf{r}) \in \mathbb{C}^{L^2 N_T N_R}$ has entries

$$[\mathbf{f}(\mathbf{r})]_{(v,k,p)} = e^{i2\pi(v\beta + k\tau + p\nu)}, v = 0, \ldots, N_T N_R - 1, \quad k, p = -N, \ldots, N.$$

Similarly as before, we use for convenience a three dimensional index to refer to entries of a vector. Finally, the matrix $\mathbf{A} \in \mathbb{C}^{N_R L \times N_R N_T L^2}$ is defined as follows. The expression

$$w_{r,p} := e^{i2\pi r N_T \beta} \sum_{j=0}^{N_T-1} e^{i2\pi j\beta} [\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}_j]_p$$

in equation (7.3) can be written as

$$w_{r,p} = \sum_{j=0}^{N_T-1} \sum_{k=-N}^{N} a_{p,k,j} e^{i2\pi(k\tau + p\nu + (j+N_T r)\beta)},$$

with

$$a_{p,k,j} = \frac{1}{L} \sum_{\ell=-N}^{N} [\mathbf{x}_j]_\ell e^{i2\pi(\ell-p)\frac{k}{L}}.$$

Let $\mathbf{f}_{p,j} \in \mathbb{C}^L$ be the vector with kth entry $[\mathbf{f}_{p,j}]_k = a_{p,k,j}$, $k = -N, \ldots, N$, and let $\mathbf{A}_j \in \mathbb{C}^{L \times L^2}$ be the block-diagonal matrix with $\mathbf{f}_{p,j}^T$ on its pth diagonal, $p = -N, \ldots, N$. With this notation, $\mathbf{A}$ is defined as the block-diagonal matrix with the matrix $[\mathbf{A}_0, \ldots, \mathbf{A}_{N_T-1}] \in \mathbb{C}^{L \times N_T L^2}$ on its diagonal, for all N_R blocks on the diagonal. With this notation, (7.3) becomes (7.20).

Similarly as for the SISO radar problem, recovery of the unknowns $b_k, \mathbf{r}_k = [\beta_k, \tau_k, \nu_k]$ from the measurement $\mathbf{z} = \sum_{k=1}^{S} b_k \mathbf{f}(\mathbf{r}_k)$ is a 3D line spectral estimation problem that can be solved with standard spectral estimation techniques. In order to recover the vector $\mathbf{z}$ from the measurement $\mathbf{y}$, we use that $\mathbf{z}$ is a sparse linear combination of atoms in the set $\mathcal{A} := \{\mathbf{f}(\mathbf{r}), \mathbf{r} \in [0,1]^3\}$, and estimate $\mathbf{z}$ by solving the basis pursuit type atomic norm minimization problem problem

$$\text{AN}(\mathbf{y}): \quad \underset{\tilde{\mathbf{z}}}{\text{minimize}} \, \|\tilde{\mathbf{z}}\|_\mathcal{A} \text{ subject to } \mathbf{y} = \mathbf{A}\tilde{\mathbf{z}},$$

where

$$\|\mathbf{z}\|_\mathcal{A} := \inf_{b_k \in \mathbb{C}, \mathbf{r}_k \in [0,1]^3} \left\{ \sum_k |b_k| : \mathbf{z} = \sum_k b_k \mathbf{f}(\mathbf{r}_k) \right\}.$$

To summarize, as for the SISO radar problem, we estimate the parameters $b_k, \mathbf{r}_k$ from $\mathbf{y}$ by:

i. solving $\text{AN}(\mathbf{y})$ in order to obtain $\mathbf{z}$,
ii. estimating the $\mathbf{r}_k$ from $\mathbf{z}$ by solving the corresponding 3D-line spectral estimation problem, and
iii. solving the linear system of equations $\mathbf{y} = \sum_{k=0}^{S-1} b_k \mathbf{A}\mathbf{f}(\mathbf{r}_k)$ for the b_k.

7.6.4 Recovery guarantees for MIMO atomic norm minimization

As before, we take the probing signals to be random by choosing its samples, i.e., the entries of the $\mathbf{x}_j$ as i.i.d. Gaussian (or sub-Gaussian) zero-mean random vari-

ables with variance $1/(N_T L)$. Moreover, we again require a minimum separation condition to be satisfied.

DEFINITION 7.5 (MIMO minimum separation condition)　We say the triplets $(\beta_j, \tau_j, \nu_j) \in [0,1]^2, j = 1, \ldots, S$, satisfy the minimum separation condition if for all $j, j' \colon j \neq j'$,

$$|\beta_j - \beta_{j'}| \geq \frac{10}{N_T N_R - 1} \quad \text{or} \quad |\tau_j - \tau_{j'}| \geq \frac{5}{N} \quad \text{or} \quad |\nu_j - \nu_{j'}| \geq \frac{5}{N}. \quad (7.21)$$

As before, $|\tau_j - \tau_{j'}|$ is the wrap-around distance on the unit circle.

Note that the triplets must *not* be separated in angle, time, *and* frequency simultaneously, for the MIMO minimum separation condition to be satisfied, it is sufficient if they are separated in at least one of those dimensions.

THEOREM 7.6　*Assume that the samples of the probing signals $\mathbf{x}_j, j = 0, \ldots, N_T - 1$, are i.i.d. zero-mean Gaussian random variables with variance $1/(N_T L)$, and let $L = 2N + 1 \geq 1024$ and $N_T N_R \geq 1024$. Consider a signal where the signs of the attenuation factors $\{b_j\}_{j=1}^S$ are i.i.d. uniform on $\{-1, 1\}$ or the complex unit disc, and suppose that the triplets $\{(\beta_j, \tau_j, \nu_j)\}_{j=1}^S$ obey the MIMO minimum separation condition. Furthermore, choose $\delta > 0$ and assume that the number of non-zero attenuation factors, S, obeys*

$$S \leq c \frac{\min(L, N_T N_R)}{\log^3 (L/\delta)}, \quad (7.22)$$

where c is a numerical constant. Then, with probability at least $1 - \delta$, $\mathbf{z} = \sum_{k=1}^S b_k \mathbf{f}(\mathbf{r}_k)$ is the unique minimizer of $\mathrm{AN}(\mathbf{y})$, $\mathbf{y} = \mathbf{A}\mathbf{z}$.

Theorem 7.2 guarantees that, with high probability, the attenuation factors and location parameters can be recovered perfectly from the observation $\mathbf{y}$ by solving a convex program (recall that the parameters $b_k, \mathbf{r}_k$ can be obtained from $\mathbf{z}$), provided that the locations $\mathbf{r}_k = [\beta_k, \tau_k, \nu_k]$ are sufficiently separated in *either* angle, time, or frequency, and provided that the total number of objects satisfies condition (7.22). Note that, translated to the physical parameters $\bar{\tau}_k, \bar{\nu}_k$, the MIMO minimum separation condition becomes: For all $k, k' \colon k \neq k'$,

$$|\beta_k - \beta_{k'}| \geq \frac{10}{N_T N_R - 1} \quad \text{or} \quad |\bar{\tau}_k - \bar{\tau}_{k'}| \geq \frac{10.01}{B} \quad \text{or} \quad |\bar{\nu}_k - \bar{\nu}_{k'}| \geq \frac{10.01}{T}.$$

Theorem 7.2 is essentially optimal in the number of objects that can be located, since S can be linear—up to a log-factor—in $\min(L, N_T N_R)$, and $S \leq \min(L, N_T N_R)$ is a necessary condition to uniquely recover the attenuation factors b_k even if the locations $\mathbf{r}_k$ are known. To see this, note that for the linear system of equations (7.20) to have a unique solution, the vectors $\mathbf{A}\mathbf{f}(\mathbf{r}_k)$ must be linearly independent. If $\beta_k = 0$, for all k, or if $\tau_k = 0$ and $\nu_k = 0$, for all k, the vectors $\mathbf{A}\mathbf{f}(\mathbf{r}_k), \mathbf{r}_k = (\beta_k, \tau_k, \nu_k), k = 0, \ldots, S - 1$ can only be linearly independent provided that $S \leq L$ and $S \leq N_T N_R$, respectively. This is seen

from

$$\mathbf{Af}(\mathbf{r}) = \begin{bmatrix} e^{i2\pi 0\beta} \sum_{j=0}^{N_T-1} e^{i2\pi j\beta} \mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}_j \\ \vdots \\ e^{i2\pi N_T(N_R-1)\beta} \sum_{j=0}^{N_T-1} e^{i2\pi j\beta} \mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}_j \end{bmatrix}.$$

We finally note that Theorem 7.2 is proven by constructing a dual certificate in a similar manner as we constructed the certificate for the SISO in Section 7.7.

7.6.5   MIMO super-resolution radar on a fine grid

As discussed before for the SISO radar setup, a practical approach to estimate the parameters $\mathbf{r}_k$ from the received signals, is to suppose the angle-time-frequency triplets lie on a *fine* grid, and solve the recovery problem on that grid. In general this leads to a gridding error, that, however, decreases as the grid becomes finer. We next discuss the corresponding (discrete) sparse signal recovery problem.

Suppose the parameters (β_k, τ_k, ν_k) lie on a grid with spacing $(1/K_1, 1/K_2, 1/K_3)$, where K_1, K_2, K_3 are integers obeying $K_1 \geq N_T N_R$, $K_2, K_3 \geq L = 2N+1$. With this assumption, the super-resolution MIMO radar problem reduces to the recovery of the sparse (discrete) signal $\mathbf{b} \in \mathbb{C}^{K_1 K_2 K_3}$ from the measurement

$$\mathbf{y} = \mathbf{Rb},$$

where $\mathbf{R} \in \mathbb{C}^{N_R L \times K_1 K_2 K_3}$ is the matrix with (n_1, n_2, n_3)-th column given by

$$\mathbf{Af}(\mathbf{r_n}), \quad \mathbf{r_n} = (n_1/K_1, n_2/K_2, n_3/K_3).$$

Note that the non-zeros of $\mathbf{b}$ and its indices correspond to the attenuation factors b_k and the locations $\mathbf{r}_k$ on the grid. A standard approach to the recovery of the sparse signal $\mathbf{b}$ from the underdetermined linear system of equations $\mathbf{y} = \mathbf{Rb}$ is to solve the following convex program:

$$\mathrm{L1}(\mathbf{y})\colon \quad \underset{\tilde{\mathbf{b}}}{\mathrm{minimize}} \left\| \tilde{\mathbf{b}} \right\|_1 \text{ subject to } \mathbf{y} = \mathbf{R}\tilde{\mathbf{b}}. \tag{7.23}$$

Below is the main result for recovery on the fine grid.

THEOREM 7.7   *Assume $L = 2N + 1 \geq 1024$, $N_T N_R \geq 1024$, and suppose we observe $\mathbf{y} = \mathbf{Rb}$, where $\mathbf{b}$ is a sparse vector with non-zeros indexed by the support set $\mathcal{S} \subseteq [K_1] \times [K_2] \times [K_3]$, $[K] := \{0, \ldots, K - 1\}$. Suppose that those indices satisfy the following minimum separation condition: For all triplets $(n_1, n_2, n_3), (n_1', n_2', n_3') \in \mathcal{S}$,*

$$\frac{|n_1 - n_1'|}{K_1} \geq \frac{10}{N_T N_R - 1} \quad or \quad \frac{|n_2 - n_2'|}{K_2} \geq \frac{5}{N} \quad or \quad \frac{|n_3 - n_3'|}{K_3} \geq \frac{5}{N}.$$

Moreover, we assume that the signs of the non-zeros of $\mathbf{b}$ are chosen independently from symmetric distributions on the complex unit circle. Choose $\delta > 0$ and assume

$$S \leq c \frac{\min(L, N_T N_R)}{\log^3(L/\delta)},$$

where c is a numerical constant. Then, with probability at least $1 - \delta$, $\mathbf{s}$ is the unique minimizer of $\mathrm{L1}(\mathbf{y})$ *in (7.23).*

Note that Theorem 7.7 does not impose any restriction on K_1, K_2, K_3, in particular they can be arbitrarily large. The proof of Theorem 7.7 is closely linked to that of Theorem 7.6; specifically, similarly as for the SISO case, the existence of a certain dual certificate guarantees that $\mathbf{b}$ is the unique minimizer of $\mathrm{L1}(\mathbf{y})$. The dual certificate is obtained directly from the dual certificate for the continuous case, which, as mentioned before, is constructed for the MIMO case in a similar way as the certificate for the SISO case has been constructed in Section 7.7.

7.6.6   Numerical results and robustness

Paralleling the discussion in Section 7.4.2 for SISO radar, we next briefly numerically evaluate the resolution obtained by the norm minimization approach to enable super-resolution in a MIMO radar, and demonstrate robustness to noise. We discuss a synthetic experiment from [15]. In that experiment, a problem instance was generated by setting $N_T = 3, N_R = 3, L = 41$, and $S = 5$ object locations (β_k, τ_k, ν_k) were drawn uniformly at random from $[0,1] \times [0, 2/\sqrt{L}]^2$. Moreover, we choose $K_1 = \mathrm{SRF} N_T N_R, K_2 = \mathrm{SRF} L$, and $K_3 = \mathrm{SRF} L$, where $\mathrm{SRF} \geq 1$ can be interpreted as a super-resolution factor as it determines by how much the $(1/K_1, 1/K_2, 1/K_3)$ grid is finer than the original, coarse grid $(1/(N_T N_R), 1/L, 1/L)$. To account for additive noise, as before, we solve the following modification of $\mathrm{L1}(\mathbf{y})$ in (7.23)

$$\mathrm{L1\text{-}ERR}\colon \underset{\tilde{\mathbf{b}}}{\mathrm{minimize}} \left\| \tilde{\mathbf{b}} \right\|_1 \text{ subject to } \left\| \mathbf{y} - \mathbf{R}\tilde{\mathbf{b}} \right\|_2^2 \leq \delta,$$

with δ chosen on the order of the noise variance. There are two error sources incurred by this approach: the gridding error obtained by assuming the points lie on a grid with spacing $(1/K_1, 1/K_2, 1/K_3)$, which decreases in SRF and becomes negligible, and the additive noise error, which is constant. The results of the simulations, depicted in Figure 7.6, show that the object resolution of the super-resolution approach is significantly better than that of the compressed sensing based approach [7, 8] corresponding to recovery on the coarse grid, i.e., SRF = 1. Moreover, the results show that the approach is robust to noise and that even under noise, the localization accuracy is significantly improved over a standard approach to radar.

We next compare our approach to the Iterative Adaptive Approach (IAA) [36], proposed for MIMO radar in the paper [37]. IAA is based on weighted least squares and has been proposed in the array processing literature. IAA can work well even with one snapshot only and can therefore be directly applied to the MIMO super-resolution problem. However, to the best of our knowledge, no analytical performance guarantees are available in the literature that attest IAA similar performance than the ℓ_1-minimization based approach. We compare the
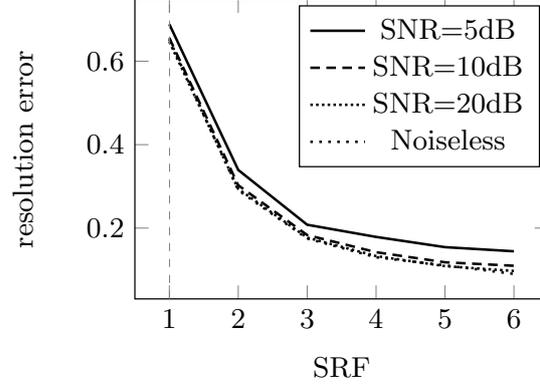
Figure 7.6 Resolution error for the recovery of $S = 5$ objects from the samples $\mathbf{y}$ with and without additive Gaussian noise $\mathbf{n}$ of a certain signal-to-noise ratio $\mathrm{SNR} = \|\mathbf{y}\|_2^2 / \|\mathbf{n}\|_2^2$, for varying super-resolution factors (SRFs). The resolution error is defined as the average over $(N_T^2 N_R^2 (\hat{\beta}_k - \beta_k)^2 + L^2 (\hat{\tau}_k - \tau_k)^2 + L^2 (\hat{\nu}_k - \nu_k)^2)^{1/2}$, $k = 1, \ldots, S$, where $(\hat{\beta}_k, \hat{\tau}_k, \hat{\nu}_k)$ are the locations obtained by solving L1-ERR.

IAA algorithm [36, Table II, "The IAA-APES Algorithm"] to L1-ERR, for a problem with parameters $N_T = 3, N_R = 3$, and $L = 41$, as before, but with $\mathrm{SRF} = 3$ and $(\beta_k, \tau_k, \nu_k) = (k/(N_R N_t), k/L, k/L), k = 1, \ldots, S$, so that the location parameters lie on the fine grid, and are separated. As before, we draw the corresponding attenuation factors b_k i.i.d. uniformly at random from the complex unit disc. Our results, depicted in Figure 7.7, show that L1-ERR performs better in this experiment than IAA, in particular for small signal-to-noise ratios.
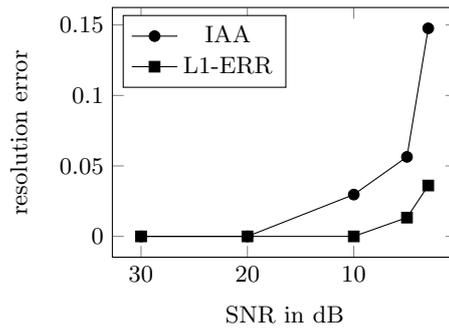

Figure 7.7 Resolution error (smaller is better) of L1-ERR and IAA applied to $\mathbf{y} + \mathbf{n}$, where $\mathbf{n} \in \mathbb{C}^{N_R L}$ is additive Gaussian noise, such that the signal-to-noise ratio is $\mathrm{SNR} := \|\mathbf{y}\|_2^2 / \|\mathbf{n}\|_2^2$. As before, the resolution error is defined as $(N_T^2 N_R^2 (\hat{\beta}_k - \beta_k)^2 + L^2 (\hat{\tau}_k - \tau_k)^2 + L^2 (\hat{\nu}_k - \nu_k)^2)^{1/2}$, where $(\hat{\beta}_k, \hat{\tau}_k, \hat{\nu}_k)$ are the locations obtained by solving L1-ERR.

7.7 Discussion and current and future research directions

In this section we discuss a class of signal recovery problems that are closely related to the SISO and MIMO radar problem, corresponding results and open theoretical research problems, and comment on computational challenges in applying the methods discussed here in practical radar systems.

The SISO and MIMO radar problems discussed in this chapter are versions of a more general problem, namely that of recovering a signal that is sparse in a *continuously* indexed dictionary, with the index corresponding to the locations and velocities of objects. In contrast, traditional compressive sensing research has focused on the recovery of signals that are sparse in *discretely* indexed dictionaries via convex programs [3] amongst other methods. As discussed in this chapter in the context of radar, signals that are sparse in continuously indexed dictionaries can be recovered via a convex program either by solving an atomic norm minimization problem, or by discretizing the continuous parameter space. However, the discretization step induces a gridding error. While in practice—provided the grid is chosen sufficiently fine—the gridding error is negligible, fine discretization leads to dictionaries with extremely correlated, i.e., coherent, columns, and the theory of compressive sensing, and many practical algorithms, rely on the dictionary to be incoherent and therefore does not apply to fine grids. The primary difficulty with recovering signals in such dictionaries is that the elements of the dictionaries are very close to each other—which is the case both for continuously indexed dictionaries as well as for finely discretized signals. Stable recovery of signals that are sparse in such dictionaries requires excluding signals that are supported on elements of the dictionary that are very close to each other. Such signals are excluded here by imposing the minimum separation condition.

More specifically, the SISO and MIMO radar problems belong to a class of signal recovery problems where the goal is to recover unknown coefficients $\{b_j\}$ and location parameters $\{\mathbf{r}_j\}$ from the measurement

$$\mathbf{y} = \sum_{j=1}^{S} b_j \mathbf{A} \mathbf{f}(\mathbf{r}_j).$$

Here, $\mathbf{f}(\mathbf{r})$ is a vector containing complex exponentials, and $\mathbf{r}$ is a d-dimensional location parameter. For example, if $\mathbf{r}$ is a one-dimensional location parameter then $[\mathbf{f}(\mathbf{r})]_r = e^{-i2\pi r\tau}$, and if $\mathbf{r}$ is two-dimensional location vector, as in the SISO radar problem, then $[\mathbf{f}(\mathbf{r})]_{(r,q)} = e^{-i2\pi(r\tau+q\nu)}$. The matrix $\mathbf{A}$ is a problem dependent matrix that parameterizes the dictionary; in the SISO case it is equal to $\mathbf{A} = \mathbf{G}_{\mathbf{x}}\mathbf{F}^H$, see equation (7.7). In other words, radar signals are sparse in a continuously dictionary that is parameterized by a matrix $\mathbf{A}$. We hasten to add that there are a number of interesting signal recovery problems in continuously indexed dictionaries that are not of this particular form, the deconvolution problem in the paper [38] is such an example, and the computational imaging problem in [39] is another.

7.7.1   Stability to noise

In this section, we discuss analytical results on the stability to noise of the atomic norm minimization framework. While currently there are no formal results for the SISO and MIMO radar problem, we discuss statements pertaining to two closely related problems. The first is the classical line spectral estimation problem where $\mathbf{A}$ is the identity matrix, and the second is a generalized line spectral estimation problem, where $\mathbf{A}$ is a Gaussian random matrix.

As mentioned before, for $\mathbf{A} = \mathbf{I}$, the sparse recovery problem reduces to the classical line spectral estimation problem studied for the noisy and noiseless case in [10, 40, 41, 12]. This problem is well understood, and atomic norm minimization succeeds under very general conditions. Specifically, the paper [10] shows that a convex program provably recovers the coefficients $\{b_j\}$ and location parameters $\{\mathbf{r}_j\}$ perfectly, provided the minimum separation condition holds (see Section 7.3.2). While standard spectral estimation techniques such as Prony's method, MUSIC, and ESPRIT [9] also provably succeed for the noiseless case, even without requiring a separation condition, an advantage of the convex program is that it does not require knowledge of S, and perhaps more importantly is provably robust [42, 40, 12].

Specifically, Tang et al. [42] show that the atomic norm regularized least squares estimate enables near-optimal denoising of $\mathbf{z}$ from a noisy measurement $\mathbf{y} = \mathbf{z} + \mathbf{e}$, where $\mathbf{z}$ is a signal of the form $\mathbf{z} = \sum_{j=1}^{S} \mathbf{f}(\nu_j)$, and $\mathbf{e}$ is zero-mean Gaussian noise with variance $\sigma^2 \mathbf{I}$. Specifically, provided the minimum separation condition holds, the atomic norm regularized least squares estimate $\hat{\mathbf{z}}$ obeys

$$\|\hat{\mathbf{z}} - \mathbf{z}\|_2^2 \leq c\sigma^2 S \log(L),$$

with high probability. This result is essentially optimal; to see this, note that even if we knew the location parameters $\{\nu_j\}$ exactly, the best bound we could achieve would be $\sigma^2 S$ [43], only by a logarithmic factor short of the result reviewed above. In addition, the paper [42] shows that the corresponding estimator localizes the frequencies up to a certain (small) error, provided that the number of samples is sufficiently large.

Next, suppose that $\mathbf{A}$ is a $M \times L$ Gaussian random matrix with i.i.d. $\mathcal{N}(0, 1/M) + i\mathcal{N}(0, 1/M)$ entires, with M typically much smaller than L. Assume we are given a noisy measurement $\mathbf{y} = \mathbf{A}\mathbf{z} + \mathbf{e}$, with $\mathbf{z}$ a signal of the form $\mathbf{z} = \sum_{j=1}^{S} \mathbf{f}(\nu_j)$ and $\mathbf{e}$ is noise. To estimate the signal $\mathbf{z}$ from such a measurement one can use an atomic norm optimization problem of the form

$$\hat{\mathbf{z}} := \arg\min_{\bar{\mathbf{z}}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\bar{\mathbf{z}}\|_2^2 \quad \text{subject to} \quad \|\bar{\mathbf{z}}\|_{\mathcal{A}} \leq \tau, \qquad (7.24)$$

with τ a tuning parameter. Theorem 4 in the paper [16] shows that, as long as

$$M \geq cS \log(L),$$

with c a fixed numerical constant, the minimizer of (7.24) with $\tau = \|\mathbf{z}\|_\mathcal{A}$ obeys

$$\|\hat{\mathbf{z}} - \mathbf{z}\|_2^2 \leq cS \log(L)\sigma^2,$$

with high probability. Again, this results is essentially optimal both with respect to the number of measurements M required relative to the number of unknowns S, as well as with respect to the best bound we can achieve for estimation under noise. Moreover, this results do not make any assumptions on the coefficients b_j. Unfortunately, the corresponding proof strategy does not carry over to the case where $\mathbf{A}$ is a structured random matrix, as in the SISO and MIMO radar problem considered here. However, numerical simulations suggest that for a number of structured random matrices, including the matrices parametrizing the SISO and MIMO radar problems, the performance of the nuclear norm minimization program as well as ℓ_1-norm regularized least squares, is similar.

7.7.2 Computational challenges

A challenge in applying the convex optimization based super-resolution methods in the context of radar is their computational complexity. Specifically, in the context of SISO radar, if we estimate the time-frequency shifts based on solving (the dual of) the atomic norm minimization problem, then the optimization variable of the corresponding convex program has dimensions $L^2 \times L^2$. Thus, an algorithm that solves or approximates the atomic norm minimization problem has computational complexity at least L^4, which is infeasible for real-world problems. As discussed in Section 7.4, what comes to our rescue is that in practice, we can solve the super-resolution radar problem on a fine grid, and recover the signal by solving a ℓ_1-minimization problem. The complexity of numerically solving the corresponding program with a standard iterative algorithm such as FISTA [44] depends on the dimension of the matrix (determined by the problem size (BT and number of antennas)), as well as the corresponding super-resolution factor (see Section 7.4.1), and the conditioning of the matrices involved. Increasing the super-resolution factor leads to both a larger problem size (the number of columns in the SISO radar problem increases quadratically in the super-resolution factor), which results in a larger iteration complexity, as well as in general to a slower convergence of the iterative algorithm, since the conditioning of the matrices involved become worse. Thus, an interesting research direction is to develop computationally efficient algorithms for the recovery of signals in continuously indexed dictionaries in general, and for the SISO and MIMO radar problem in particular. See the papers [45, 46] for some recent work in this direction.

References

[1] D. W. Bliss and K. W. Forsythe, "Multiple-input multiple-output (MIMO) radar and imaging," in *Asilomar Conf. on Signals, Syst. and Comput.*, 2003, pp. 54–59.

[2] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 106–114, 2007.

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

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

[5] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *IEEE Radar Conf.*, 2007, pp. 128–133.

[6] R. Heckel and H. Bölcskei, "Identification of sparse linear operators," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7985–8000, 2013.

[7] D. Dorsch and H. Rauhut, "Refined analysis of sparse MIMO radar," *J. Fourier Anal. Appl.*, vol. 23, 2017.

[8] T. Strohmer and H. Wang, "Adventures in compressive sensing based MIMO Radar," in *Excursions in Harm. Anal.*, ser. Appl. Num. Harm. Anal., 2015, pp. 285–326.

[9] P. Stoica and R. L. Moses, *Spectral Analysis of Signals.* Pearson Prentice Hall, 2005.

[10] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Comm. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014.

[11] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[12] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, 2013.

[13] C. Aubel, D. Stotz, and H. Bölcskei, "A theory of super-resolution from short-time Fourier transform measurements," *J. Fourier Anal. Appl.*, vol. 24, 2018.

[14] R. Heckel, V. I. Morgenshtern, and M. Soltanolkotabi, "Super-resolution radar," *Inf. Inference*, vol. 5, no. 1, pp. 22–75, 2016.

[15] R. Heckel, "Super-resolution mimo radar," in *IEEE International Symposium on Information Theory*, 2016, pp. 1416–1420.

[16] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," *IEEE Trans. Inf. Theory*, vol. 64, pp. 4001–4023, 2017.

[17] T. Strohmer, "Pseudodifferential operators and Banach algebras in mobile communications," *Appl. Comput. Harmon. Anal.*, vol. 20, no. 2, pp. 237–249, 2006.

[18] G. Tauböck, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, 2010.

[19] W. Bajwa, A. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. of 46th Allerton Conf. on Commun., Control, and Comput.*, Monticello, IL, 2008, pp. 575–582.

[20] W. U. Bajwa, K. Gedalyahu, and Y. C. Eldar, "Identification of parametric underspread linear systems and super-resolution radar," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2548–2561, 2011.

[21] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, no. 3, pp. 292–300, 1976.

[22] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Commun. Pur. Appl. Math.*, vol. 67, no. 11, pp. 1877–1904, 2014.

[23] A. Gershman and N. Sidiropoulos, Eds., *Space-Time Processing for MIMO Communications.* Chichester; Hoboken, NJ: Wiley, 2005.

[24] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.

[25] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution," *IEEE Trans. Inform. Theory*, vol. 62, 2015.

[26] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, no. 4, pp. 360–393, 1963.

[27] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM J. on Math. Anal.*, vol. 23, no. 5, pp. 1309–1331, 1992.

[28] A. Moitra, "Super-resolution, extremal functions and the condition number of vandermonde matrices," in *Proc. of the Forty-seventh Annual ACM Symposium on Theory of Computing*, 2015, pp. 821–830.

[29] A. Quinquis, E. Radoi, and F. C. Totir, "Some radar imagery results using super-resolution techniques," *IEEE Trans. Antennas Propag.*, vol. 52, no. 5, pp. 1230–1244, 2004.

[30] A. Jakobsson, A. L. Swindlehurst, and P. Stoica, "Subspace-based estimation of time delays and Doppler shifts," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2472–2483, 1998.

[31] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries-just discretize," in *Asilomar Conf. on Signals, Syst. and Comput.*, Pacific Grove, CA, 2013, pp. 1043–1047.

[32] S. R. Becker, E. J. Candès, and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Prog. Comp.*, vol. 3, no. 3, pp. 165–218, 2011.

[33] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[34] B. Friedlander, "On the relationship between MIMO and SIMO radars," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 394–398, Jan. 2009.

[35] T. Strohmer and B. Friedlander, "Analysis of sparse MIMO radar," *Appl. Comput. Harm. Anal.*, vol. 37, no. 3, pp. 361–388, 2014.

[36] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 425–443, 2010.

[37] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. Sadjadi, "Iterative adaptive approaches to MIMO radar imaging," *IEEE J. Sel. Topics Signal Process*, vol. 4, no. 1, pp. 5–20, Feb. 2010.

[38] B. Bernstein and C. Fernandez-Granda, "Deconvolution of point sources: A sampling theorem and robustness guarantees," *Commun. Pur. Appl. Math., to appear*, 2017.

[39] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," *Optica*, vol. 5, no. 1, p. 19, 2018.

[40] E. Candes and C. Fernandez-Granda, "Super-Resolution from Noisy Data," *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229–1254, 2014.

[41] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Information and Inference*, vol. 5, pp. 251–303, 2016.

[42] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499–512, 2015.

[43] F. Bunea, A. Tsybakov, and M. Wegkamp, "Sparsity oracle inequalities for the Lasso," *Electronic Journal of Statistics*, vol. 1, pp. 169–194, 2007.

[44] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, p. 183202, 2009.

[45] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[46] N. Rao, P. Shah, and S. Wright, "Forward-backward greedy algorithms for atomic norm regularization," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5798–5811, Nov. 2015.