Convex recovery from interferometric measurements

Laurent Demanet and Vincent Jugnon

Department of Mathematics, MIT

Abstract—This paper discusses some questions that arise when a linear inverse problem involving $Ax = b$ is reformulated in the interferometric framework, where quadratic combinations of b are considered as data in place of b.

First, we show a deterministic recovery result for vectors x from measurements of the form $(Ax)_i \overline{(Ax)_j}$ for some left-invertible A. Recovery is exact, or stable in the noisy case, when the couples (i, j) are chosen as edges of a well-connected graph. One possible way of obtaining the solution is as a feasible point of a simple semidefinite program. Furthermore, we show how the proportionality constant in the error estimate depends on the spectral gap of a data-weighted graph Laplacian.

Second, we present a new application of this formulation to interferometric waveform inversion, where products of the form $(Ax)_i \overline{(Ax)_j}$ in frequency encode generalized cross-correlations in time. We present numerical evidence that interferometric *inversion* does not suffer from the loss of resolution generally associated with interferometric imaging, and can provide added robustness with respect to specific kinds of kinematic uncertainties in the forward model A.

I. INTRODUCTION

Throughout this paper, we consider complex quadratic measurements of $x \in \mathbb{C}^n$ of the form

$$B_{ij} = (Ax)_i \overline{(Ax)_j}, \qquad (i, j) \in E, \qquad (1)$$

for certain well-chosen couples of indices (i, j), a scenario that we qualify as "interferometric". This combination is special in that it is symmetric in x, and of rank 1 with respect to the indices i and j.

The regime that interests us is when the number m of measurements, i.e., of couples (i, j) in E, is comparable to the number n of unknowns. While phaseless measurements $b_i = |(Ax)_i|^2$ only admit recovery when A

has very special structure – such as, being a tall random matrix with Gaussian i.i.d. entries [7], [11] – products $(Ax)_i \overline{(Ax)_j}$ for $i \neq j$ correspond to the idea of phase differences, hence encode much more information. As a consequence, stable recovery occurs under very general conditions: left-invertibility of A and "connectedness" of set E of couples (i, j). These conditions suffices to allow for m to be on the order of n. Various algorithms return x accurately up to a global phase; we mosty discuss variants of lifting with semidefinite relaxation in this paper. In contrast to other recovery results in matrix completion [8], [24], *no randomness is needed in the data model*, and our proof technique involves elementary spectral graph theory rather than dual certification or uncertainty principles.

The mathematical content of this paper is the formulation of a quadratic analogue of the well-known relative error bound

$$\frac{\|x - x_0\|}{\|x_0\|} \leq \kappa(A) \frac{\|e\|}{\|b\|} \qquad (2)$$

for the least-squares solution of the overdetermined linear system $Ax = b$ with $b = Ax_0 + e$, and where $\kappa(A)$ is the condition number of A. Our result is that an inequality of the form (2) still holds in the quadratic case, but with the *square root of the spectral gap of a data-weighted graph Laplacian* in place of $\|b\|$ in the right-hand side. This spectral gap quantifies the connectedness of E, and has the proper homogeneity with respect to b.

The numerical results mostly concern the case when A is a forward model that involves solving a linearized fixed-frequency wave equation, as in seismic or radar imaging. In case x is a reflectivity, Ax is the wavefield that results from an incoming wave being scattered by x, and $(Ax)_i \overline{(Ax)_j}$ has a meaning in the context of interferometry, as we explain in the next section. Our claims are that

- Stable recovery holds, and the choice of measurement set E for which this is the case matches the theory in this paper;
- Interferometric inversion yields no apparent loss of resolution when compared against the classical imaging methods; and

- Some degree of robustness to specific model inaccuracies is observed in practice, although it is not explained by the theory in this paper.

A. Physical context: interferometry

In optical imaging, an interference fringe of two (possibly complex-valued) wavefields $f(t)$ and $g(t)$, where t is either a time or a space variable, is any combination of the form $|f(t) + g(t + t')|^2$. The sum is a result of the linearity of amplitudes in the fundamental equations of physics (such as Maxwell or Schrödinger), while the modulus squared is simply the result of a detector measuring intensities. The cross term $2\Re(f(t)\overline{g(t + t')})$ in the expansion of the squared modulus manifestly carries the information of destructive vs. constructive interference, hence is a continuous version of what we referred to earlier as an "interferometric measurement".

In particular, when the two signals are sinusoidal at the same frequency, the interferometric combination highlights a phase difference. In astronomical interferometry, the delay t' is for instance chosen so that the two signals interfere constructively, yielding better resolution. Interferometric synthetic aperture radar (InSAR) is a remote sensing technique that uses the fringe from two datasets taken at different times to infer small displacements. In X-ray ptychograhy [25], imaging is done by undoing the interferometric combinations that the diffracted X-rays undergo from encoding masks. These are but three examples in a long list of applications.

Interferometry is also playing an increasingly important role in geophysical imaging, i.e., inversion of the elastic parameters of the portions of the Earth's upper crust from scattered seismic waves. In this context however, the signals are more often impulsive than monochromatic, and interferometry is done as part of the *computational processing rather than the physical measurements*[1]. An interesting combination of two seismogram traces f and g at nearby receivers is then $\widehat{f}(\omega)\overline{\widehat{g}(\omega)}$, i.e., the Fourier transform of their cross-correlation. It highlights a time lag in the case when f and g are impulses. More generally, it will be important to also consider the cross-ambiguities $\widehat{f}(\omega)\overline{\widehat{g}(\omega')}$ where $\omega' \simeq \omega$.

Cross-correlations have been shown to play an important role in geophysical imaging, mostly because of their stability to statistical fluctuations of a scattering medium [3] or an incoherent source [14], [20]. Though seismic interferometry is a vast research area, both in the exploration and global contexts [4], [13], [16], [28], [29], [32], [35], explicit inversion of reflectivity parameters from interferometric data has to our knowledge only been

considered in [12], [18]. Interferometric inversion offers great promise for model-robust imaging, i.e., recovery of reflectivity maps in a less-sensitive way on specific kinds of errors in the forward model.

Finally, interferometric measurements also play an important role in quantum optical imaging. See [27] for a nice solution to the inverse problem of recovering a scattering dielectric susceptibility from measurements of two-point correlation functions relative to two-photon entangled states.

As we revise this paper, we also note the recent success of interferometric inversion for passive synthetic-aperture radar (SAR) imaging, under the name low-rank matrix recovery [23].

B. Mathematical context and related work

The setting of this paper is discrete, hence we let i and j in place of either a time of frequency variable. We also specialize to $f = g$, and we let $f = Ax$ to possibly allow an explanation of the signal f by a linear forward model[2] A.

The link between products of the form $f\,\overline{g}$ and squared measurements $|f + g|^2$ goes both ways, as shown by the polarization identity

$$f_i\,\overline{f_j} = \frac{1}{4}\sum_{k=1}^{4} e^{-i\pi k/2}|f_i + e^{i\pi k/2}f_j|^2.$$

Hence any result of robust recovery of f, or x, from couples $f_i\,\overline{f_j}$, implies the same result for recovery from phaseless measurements of the form $|f_i + e^{i\pi k/2}f_j|^2$. This latter setting was precisely considerered by Candès et al. in [7], where an application to X-ray diffraction imaging with a specific choice of masks is discussed. In [1], Alexeev et al. use the same polarization identity to design good measurements for phase retrieval, such that recovery is possible with $m = O(n)$.

Recovery of f_i from $f_i\overline{f_j}$ for some (i, j) when $|f_i| = 1$ (interferometric phase retrieval) can be seen a special case of the problem of angular synchronization considered by Singer [30]. There, rotation matrices R_i are to be recovered (up to a global rotation) from measurements of relative rotations $R_iR_j^{-1}$ for some (i, j). This problem has an important application to cryo-electron microscopy, where the measurements of relative rotations are further corrupted in an a priori unknown fashion (i.e., the set E is to be recovered as well). An impressive recovery result under a Bernoulli model of gross corruption, with a characterization of the critical probability, were recently obtained by Wang and Singer [33]. The spectrum of an adapted graph Laplacian plays

---

[1]Time reversal is an important exception not considered in this paper, where interferometry stems from experimental acquisition rather than processing.

[2]Such as scattering from a reflectivity profile x in the Born approximation, for which A is a wave equation Green's function. Section III covers the description of A in this context.

an important role in their analysis [2], much as it does in this paper. Singer and Cucuringu also considered the angular synchronization problem from the viewpoint of rigidity theory [31]. For the similar problem of recovery of positions from relative distance, with applications to sensor network localization, see for instance [17].

The algorithmic approach considered in this paper for solving interferometric inversion problems is to formulate them via lifting and semidefinite relaxation. This idea was considered by many groups in recent years [7], [9], [17], [30], [34], and finds its origin in theoretical computer science [15].

II. MATHEMATICAL RESULTS

A. Recovery of unknown phases

Let us start by describing the simpler problem of interferometric phase recovery, when $A = I$ and we furthermore assume $|x_i| = 1$. Given a vector $x_0 \in \mathbb{C}^n$ such that $|(x_0)_i| = 1$, a set E of pairs (i, j), and noisy interferometric data $B_{ij} = (x_0)_i \overline{(x_0)_j} + \varepsilon_{ij}$, find a vector x such that

$$|x_i| = 1, \qquad \sum_{(i,j) \in E} |x_i \overline{x_j} - B_{ij}| \leq \sigma, \qquad (3)$$

for some $\sigma > 0$. Here and below, if no heuristic is provided for σ, we may cast the problem as a minimization problem for the misfit and obtain σ a posteriori.

The choice of the elementwise ℓ_1 norm over E is arbitrary, but convenient for the analysis in the sequel[3]. We aim to find situations in which this problem has a solution x close to x_0, up to a global phase. Notice that x_0 is feasible for (3), hence a solution exists, as soon as $\sigma \geq \sum_{i,j \in E} |\varepsilon_{ij}|$.

The relaxation by *lifting* of this problem is to find X (a proxy for xx^*) such that

$$X_{ii} = 1, \qquad \sum_{(i,j) \in E} |X_{ij} - B_{ij}| \leq \sigma, \qquad X \succeq 0,$$

then let x be the top eigenvector of X with $\|x\|^2 = n$.
(4)

The notation $X \succeq 0$ means that X is symmetric and positive semi-definite. Again, the feasibility problem (4) has at least one solution ($X_0 = x_0 x_0^*$) as soon as $\sigma \geq \sum_{i,j \in E} |\varepsilon_{ij}|$.

The set E generates edges of a graph $G = (V, E)$, where the nodes in V are indexed by i. Without loss of generality, we consider E to be symmetric. By convention, G does not contain loops, i.e., the diagonal $j = i$ is not part of E. (Measurements on the diagonal are not informative for the phase recovery problem, since $|(x_0)_i|^2 = 1$.)

[3]The choice of ℓ_1 norm as "least unsquared deviation" is central in [33] for the type of outlier-robust recovery behavior documented there.

The graph Laplacian on G is

$$L_{ij} = \begin{cases} d_i & \text{if } i = j; \\ -1 & \text{if } (i, j) \in E; \\ 0 & \text{otherwise,} \end{cases}$$

where d_i is the node degree $d_i = \sum_{j:(i,j) \in E} 1$. Observe that L is symmetric and $L \succeq 0$ by Gershgorin. Denote by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ the eigenvalues of L sorted in increasing order. Then $\lambda_1 = 0$ with the constant eigenvector $v_1 = \frac{1}{\sqrt{n}}$. The second eigenvalue is zero if and only if G has two or more disconnected components. When $\lambda_2 > 0$, its value is a measure of connectedness of the graph. Note that $\lambda_n \leq 2d$ by Gershgorin again, where $d = \max_i d_i$ is the maximum degree.

Since $\lambda_1 = 0$, the second eigenvalue λ_2 is called the spectral gap. It is a central quantity in the study of expander graphs: it relates to

- the edge expansion (Cheeger constant, large if λ_2 is large);
- the degree of separation between any two nodes (small if λ_2 is large); and
- the speed of mixing of a random walk on the graph (fast if λ_2 is large).

More information about spectral graph theory can be found, e.g., in the lecture notes by Lovasz [21]. It is easy to show with interlacing theorems that adding an edge to E, or removing a node from V, both increase λ_2. The spectral gap plays an important role in the following stability result.

In the sequel, we denote the componentwise ℓ_1 norm on the set E by $\| \cdot \|_1$.

Theorem 1. *Assume $\|\varepsilon\|_1 + \sigma \leq n\lambda_2$, where λ_2 is the second eigenvalue of the graph Laplacian L on G. Any solution x of (3) or (4) obeys*

$$\|x - e^{i\alpha} x_0\| \leq 4 \sqrt{\frac{\|\varepsilon\|_1 + \sigma}{\lambda_2}},$$

for some $\alpha \in [0, 2\pi)$.

The proof is in the appendix. Manifestly, recovery is exact (up to the global phase ambiguity encoded by the parameter α, because the algorithm could return another vector multiplied by some $e^{i\alpha}$ over which there is no control.) as soon as $\varepsilon = 0$ and $\sigma = 0$, provided $\lambda_2 \neq 0$, i.e., the graph G is connected. The easiest way to construct expander graphs (graphs with large λ_2) is to set up a probabilistic model with a Bernoulli distribution for each edge in an i.i.d. fashion, a model known as the Erdős-Rényi random graph. It can be shown that such graphs have a spectral gap bounded away from zero, independently of n and with high probability, with $m = O(n \log n)$ edges.

A stronger result is available when the noise ε originates at the level of x_0, i.e., $B = x_0 x_0^* + \varepsilon$ has the form $(x_0 + e)(x_0 + e)^*$.

Corollary 2. *Assume $\varepsilon = (x_0 + e)(x_0 + e)^* - x_0 x_0^*$ and $\sigma \leq n\lambda_2$, where λ_2 is the second eigenvalue of the graph Laplacian L on G. Any solution x of (3) or (4) obeys*

$$\|x - e^{i\alpha} x_0\| \leq 4\sqrt{\frac{\sigma}{\lambda_2}} + \|e\|,$$

for some $\alpha \in [0, 2\pi)$.

Proof. Apply theorem 1 with $\varepsilon = 0$, $x_0 + e$ in place of x_0, then use the triangle inequality. $\square$

In the setting of the corollary, problem (3) always has $x = x_0 + e$ as a solution, hence is feasible even when $\sigma = 0$.

Let us briefly review the eigenvector method for interferometric recovery. In [30], Singer proposed to use the first eigenvector of the (noisy) data-weighted graph Laplacian as an estimator of the vector of phases. A similar idea appears in the work of Montanari et al. as the first step of their OptSpace algorithm [19], and in the work of Chatterjee on universal thresholding [10]. In our setting, this means defining

$$(\widetilde{\mathcal{L}})_{ij} = \begin{cases} d_i & \text{if } i = j; \\ -B_{ij} & \text{if } (i,j) \in E; \\ 0 & \text{otherwise,} \end{cases}$$

and letting $x = \widetilde{v}_1 \sqrt{n}$ where v_1 is the unit-norm eigenvector of $\widetilde{\mathcal{L}}$ with smallest eigenvalue. Denote by $\widetilde{\lambda}_1 \leq \widetilde{\lambda}_2 \leq \ldots$ the eigenvalues of $\widetilde{\mathcal{L}}$. The following result is known from [2], but we provide an elementary proof (in the appendix) for completeness of the exposition.

Theorem 3. *Assume $\|\varepsilon\| \leq \widetilde{\lambda}_2/2$. Then the result x of the eigenvector method obeys*

$$\|x - e^{i\alpha} x_0\| \leq \sqrt{2n}\, \frac{\|\varepsilon\|}{\widetilde{\lambda}_2},$$

for some $\alpha \in [0, 2\pi)$.

Alternatively, we may express the inequality in terms of λ_2, the spectral gap of the noise-free Laplacian L defined earlier, by noticing[4] that $\widetilde{\lambda}_2 \geq \lambda_2 - \|\varepsilon\|$. Both λ_2 and $\widetilde{\lambda}_2$ are computationally accessible. In the case when $|B_{ij}| = 1$, we have $\widetilde{\lambda}_1 \geq 0$, hence $\widetilde{\lambda}_2$ is (slightly) greater than the spectral gap $\widetilde{\lambda}_2 - \widetilde{\lambda}_1$ of $\widetilde{\mathcal{L}}$. Note that the $1/\widetilde{\lambda}_2$ scaling appears to be sharp in view of the numerical experiments reported in section III. The inverse square root scaling of theorem 1 is stronger in the presence of small spectral gaps, but the noise scaling is weaker in theorem 1 than in theorem 3.

[4]This owes to $\|\mathcal{L} - \widetilde{\mathcal{L}}\| \leq \|\varepsilon\|$, with $\mathcal{L} = \Lambda L \Lambda^*$ the noise-free Laplacian with phases introduced at the beginning of section 2.1.

B. Interferometric recovery

The more general version of the interferometric recovery problem is to consider a left-invertible tall matrix A, linear measurements $b = Ax_0$ for some vector x_0 (without condition on the modulus of either b_i or $(x_0)_i$), noisy interferometric measurements $B_{ij} = b_i \overline{b_j} + \varepsilon_{ij}$ for (i,j) in some set E, and find x subject to

$$\sum_{(i,j) \in E \cup D} |(Ax)_i \overline{(Ax)_j} - B_{ij}| \leq \sigma. \tag{5}$$

Notice that we now take the union of the diagonal $D = \{(i,i)\}$ with E. Without loss of generality we assume that $\varepsilon_{ij} = \overline{\varepsilon_{ji}}$, which can be achieved by symmetrizing the measurements, i.e., substituting $\frac{B_{ij} + B_{ji}}{2}$ for B_{ij}.

Since we no longer have a unit-modulus condition, the relevant notion of graph Laplacian is now data-dependent. It reads

$$(L_{|b|})_{ij} = \begin{cases} \sum_{k : (i,k) \in E} |b_k|^2 & \text{if } i = j; \\ -|b_i||b_j| & \text{if } (i,j) \in E; \\ 0 & \text{otherwise.} \end{cases}$$

The connectedness properties of the underlying graph now depend on the size of $|b_i|$: the edge (i,j) carries valuable information if and only if both $|b_i|$ and $|b_j|$ are large.

A few different recovery formulations arise naturally in the context of lifting and semidefinite relaxation.

- The basic lifted formulation is to find some X such that

$$\sum_{(i,j) \in E \cup D} |(AXA^*)_{ij} - B_{ij}| \leq \sigma, \qquad X \succeq 0,$$

then let $x = x_1 \sqrt{\eta_1}$, where (η_1, x_1) is the top eigen-pair of X. (6)

Our main result is as follows, see the appendix for the proof.

Theorem 4. *Assume $\|\varepsilon\|_1 + \sigma \leq \lambda_2/2$, where λ_2 is the second eigenvalue of the data-weighted graph Laplacian $L_{|b|}$. Any solution x of (6) obeys*

$$\frac{\|x - e^{i\alpha} x_0\|}{\|x_0\|} \leq 15\, \kappa(A)^2 \sqrt{\frac{\|\varepsilon\|_1 + \sigma}{\lambda_2}},$$

for some $\alpha \in [0, 2\pi)$, and where $\kappa(A)$ is the condition number of A.

The quadratic dependence on $\kappa(A)$ is necessary[5]. In section III, we numerically verify the inverse square root scaling in terms of λ_2.

[5]The following example shows why that is the case. For any X_0 and invertible A, the solution to $AXA^* = AX_0A^* + \varepsilon$ is $X = X_0 + A^+ \varepsilon (A^*)^+$. Let $X_0 = e_1 e_1^T$, $\varepsilon = \delta e_1 e_1^*$ for some small δ, and $A^+ = I + N e_1 e_1^T$. Then $X = (1 + \delta N^2) e_1 e_1^T$, and the square root of its leading eigenvalue is $\sqrt{\eta_1} \simeq 1 + \frac{1}{2}\delta N^2$. As a result, x is perturbation of x_0 by a quantity of magnitude $O(\delta \|A^+\|^2)$.

If the noise originates from $b+e$ rather than $bb^*+\varepsilon$, the error bound is again improved to

$$\frac{\|x - e^{i\alpha}x_0\|}{\|x_0\|} \leq 15\,\kappa(A)^2\,\sqrt{\frac{\sigma}{\lambda_2}} + \kappa(A)\,\frac{\|e\|}{\|b\|},$$

for the same reason as earlier.

- An alternative, two-step lifting formulation is to find x through Y such that

$$\sum_{(i,j)\in E\cup D} |Y_{ij} - B_{ij}| \leq \sigma, \qquad Y \succeq 0,$$

then let $x = A^+ y_1 \sqrt{\eta_1}$, where (η_1, y_1) is the top eigen-pair of Y. $\qquad\qquad$ (7)

The dependence on the condition number of A is more favorable than for the basic lifting formulation.

Theorem 5. *Assume* $\|\varepsilon\|_1 + \sigma \leq \lambda_2/2$*, where* λ_2 *is the second eigenvalue of the data-weighted graph Laplacian* $L_{|b|}$*. Any solution* x *of (5) or (7) obeys*

$$\frac{\|x - e^{i\alpha}x_0\|}{\|x_0\|} \leq 15\,\kappa(A)\,\sqrt{\frac{\|\varepsilon\|_1 + \sigma}{\lambda_2}},$$

for some $\alpha \in [0, 2\pi)$.

However, this formulation may be more computationally expensive than the one-step variant if data (b) space is much larger than model (x) space.

The quantity λ_2 is not computationally accessible in general, but it can be related to the second eigenvalue $\widetilde{\lambda}_2$ of the noisy data-weighted Laplacian,

$$\left(\widetilde{\mathcal{L}}_B\right)_{ij} = \begin{cases} \sum_{k:(i,k)\in E} B_{kk} & \text{if } i = j; \\ -B_{ij} & \text{if } (i,j) \in E; \\ 0 & \text{otherwise.} \end{cases}$$

It is straightforward to show that $\lambda_2 \geq \widetilde{\lambda}_2 - [(d+1)\|\varepsilon\|_\infty + \|\varepsilon\|]$, where $\|\cdot\|_\infty$ is the elementwise maximum on $E \cup D$, $\|\cdot\|$ is the spectral norm, and d is the maximum node degree.

C. Influence of the spectral gap

In this section, we numerically confirm the tightness of the error bound for phase recovery given by theorem 1 on toy examples ($n = 2^7$), with respect to the spectral gap. We perform the corresponding experiment for the situation of theorem 3. In order to span a wide range of spectral gaps, three types of graphs are considered:

- the cycle graph P_n which is proven to be the connected graph with the smallest spectral gap[6];
- graphs obtained by adding randomly K edges to P_n with K ranging from 1 to 50;

---

[6]As justified by the decreasing property of λ_2 under edge removal, mentioned earlier.

- Erdős-Rényi random graphs with probability ranging from 0.03 to 0.05, conditioned on connectedness (positive spectral gap).

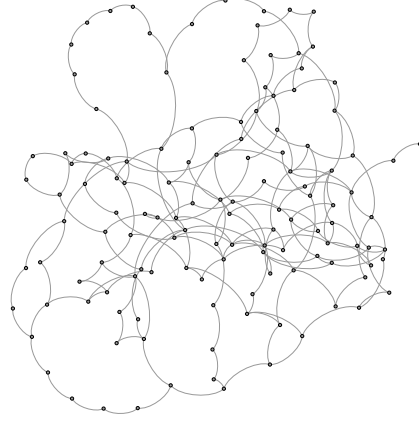A realization of the two latter types of graphs is given in figure 1.

Fig. 1. P_n + random edges (top), Erdős-Rényi random graph (bottom)

To study the eigenvector method, we draw one realization of a symmetric error matrix ε with $\varepsilon_{ij} \sim \mathbb{CN}(0, \eta^2)$, with $\eta = 10^{-8}$. The spectral norm of the noise (used in theorem 3) is $\|\varepsilon\| \sim 2 \times 10^{-7}$.

For different realizations of the aforementioned graphs, we estimate the solution with the eigenvector method

and plot the ℓ_2 recovery error as a function of $\widetilde{\lambda}_2$. See figure 2.
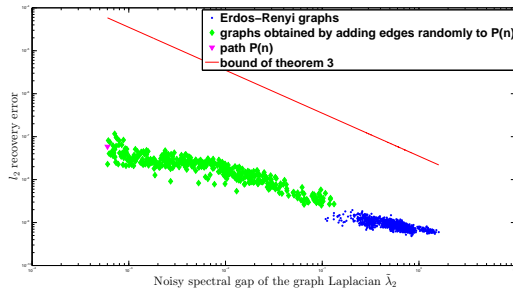

Fig. 2. Recovery error for the eigenvector method as a function of $\widetilde{\lambda}_2$

To study the feasibility method, we consider the case of an approximate fit ($\sigma = 10^{-4}$) in the noiseless case ($\varepsilon = 0$). The feasibility problem (4) is solved using the Matlab toolbox cvx which calls the toolbox SeDuMi. An interior point algorithm (centering predictor-corrector) is used. The recovery error as a function of the spectral gap λ_2 is illustrated in figure 3. The square root scaling of the theorem seems to be a good match for the numerical experiments.
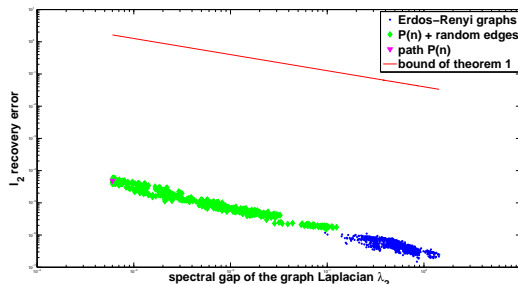

Fig. 3. Recovery error for the feasibility method as a function of λ_2

III. NUMERICAL RESULTS: INTERFEROMETRIC INVERSE SCATTERING

A. Setup and context

An important application of interferometric inversion is seismic imaging, where the (simplest) question is to recover of a medium's wave velocity map from recordings of waves scattered by that medium. Linear inverse problems often arise in this context. For instance, inverse source problems arise when locating microseismic events, and linear inverse scattering in the Born regime yield model updates for subsurface imaging. These linear problems all take the form

$$Fm = d, \qquad (8)$$

where F is the forward or modeling operator, describing the wave propagation and the acquisition, m is the reflectivity model, and d are the observed data. (We make this choice of notation from now on; it is standard and more handy than $Ax = b$ for what follows.) The classical approach is to use the data (seismograms) directly, to produce an image either

- by migrating the data (Reverse-time migration, RTM),

$$I_{\text{RTM}} = \widetilde{F}^* d$$

where $\widetilde{F}$ is the **simulation** forward operator and $*$ stands for the adjoint ;
- or by finding a model that best fits the data, in a least-squares sense (Least-squares migration, LSRTM),

$$m_{\text{LSM}} = \arg\min_m \frac{1}{2}||\widetilde{F}m - d||_2^2. \qquad (9)$$

It is important to note that the physical forward operator F and the one used in simulation $\widetilde{F}$ can be different due to modeling errors or uncertainty. Such errors can happen at different levels:

1) background velocity,
2) sources and receivers positions,
3) sources time profiles.

This list is non-exhaustive and these modeling errors have a very different effect from additive Gaussian white noise, in the sense that they induce a coherent perturbation in the data. As a result, the classical approaches (RTM, LSM) may fail in the presence of such uncertainties.

The idea of using interferometry (i.e. products of pairs of data) to make migration robust to modeling uncertainties has already been proposed in the literature [4], [26], [29], producing remarkable results. In their 2005 paper [4], Borcea et al. developed a comprehensive framework for interferometric migration, in which they proposed the Coherent INTerfermetic imaging functional (CINT), which can be recast in our notation as

$$I_{\text{CINT}} = \text{diag}\{\widetilde{F}^*(E \circ [dd^*])\widetilde{F}\},$$

where dd^* is the matrix of all data pairs products, E is a selector, that is, a sparse matrix with ones for the considered pairs and zeros elsewhere, $\circ$ is the entrywise (Hadamard) product, and diag is the operation of extracting the diagonal of a matrix. The CINT functional involves $\widetilde{F}^*$, $\widetilde{F}$ and F, F^* (implicitly through dd^*), so cancellation of model errors can still occur, even when F and $\widetilde{F}$ are different. Through a careful analysis of wave propagation in the particular case where the uncertainty consists of random fluctuations of the background velocity, Borcea et al. derive conditions on E under which CINT will be robust.

The previous sections of this paper do not inform the robustness mentioned above, but they explain *how* the power of interferometry can be extended to inversion. In

a noiseless setting, the proposal is to perform inversion using selected data pair products,

$$\text{find } m \text{ s.t. } E \circ (\widetilde{F}mm^*\widetilde{F}^*) = E \circ [dd^*], \qquad (10)$$

i.e., we look for a model m that explains the data pair products $d_i\bar{d}_j$ selected by $(i,j) \in E$. (The notation E has not changed from the previous sections.) Here, i and j are meta-indices in data space. For example, in an inverse source problem in the frequency domain, $i \equiv (r_i, \omega_i)$ and $j \equiv (r_j, \omega_j)$, and for inverse Born scattering, $i \equiv (r_i, s_i, \omega_i)$ and $j \equiv (r_j, s_j, \omega_j)$.

A straightforward and noise-aware version of this idea is to fit products in a least-squares sense,

$$\widehat{m}_{\text{LS,pairs}} = \arg\min_m ||E \circ (\widetilde{F}mm^*\widetilde{F}^* - dd^*)||_F^2. \quad (11)$$

While the problem in (10) is quadratic, the least-squares cost in (11) is quartic and nonconvex. The introduction of local minima is a highly undesired feature, and numerical experiments show that gradient descent can indeed converge to a completely wrong local minimizer.

A particular instance of the interferometric inversion problem is inversion from cross-correlograms. In that case,

$$E_{i,j} = 1 \iff \omega_i = \omega_j.$$

This means that E considers data pairs from different sources and receivers at the same frequency,

$$d_i\overline{d_j} = d(r_i, s_i, \omega)\overline{d(r_j, s_j, \omega)},$$

where the overline stands for the complex conjugation. This expression is the Fourier transform at frequency ω of the cross-correlogram between trace $(r_i.s_i)$ and trace (r_j, s_j).

The choice of selector E is an important concern. As explained earlier, it should describe a connected graph for inversion to be possible. In the extreme case where E is the identity matrix, the problem reduces to estimating the model from intensity-only (phaseless) measurements, which does not in general have a unique solution for the kind of F we consider. The same problem plagues inversion from cross-correlograms only: E does not correspond to a connected graph in that case either, and numerical inversion typically fails. On the other hand, it is undesirable to consider too many pairs, both from a computational point of view and for robustness to model errors. For instance, in the limit when E is the complete graph of all pairs (i,j), it is easy to see that the quartic cost function reduces to the square of the least-squares cost function. Hence, there is a trade-off between robustness to uncertainties and quantity of information available to ensure invertibility. It is important to stress that theorem 4 gives sufficient conditions on E for recovery to be possible and stable to additive noise ε, but not to modeling error (in the

theorem, $F = \widetilde{F}$). It does not provide an explanation of the robust behavior of interferometric inversion; see however the numerical experiments.

In the previous section, we showed that lifting convexifies problems such as (10) and (11) in a useful way. In the context of wave-based imaging, this idea was first proposed in [9] for intensity-only measurements. In this section's notations, we replace the optimization variable m by the symmetric matrix $M = mm^*$, for which the data match becomes a linear constraint. Incorporating the knowledge we have on the solution, the problem becomes

$$\text{find } M \quad \text{s. t.}$$
$$E \circ [\widetilde{F}M\widetilde{F}^*] = E \circ [dd^*],$$
$$M \succeq 0,$$
$$\text{rank}(M) = 1.$$

The first two constraints (data fit and positive semi-definiteness) are convex, but the rank constraint is not and would in principle lead to a combinatorially hard problem. However, as the theoretical results of this paper make clear, the rank constraint can often be dropped. We also relax the data pairs fit – an exact fit is ill-advised because of noise and modeling errors – to obtain the following feasibility problem equivalent to (6),

$$\text{find } M \quad \text{s. t. },$$
$$||\widetilde{F}M\widetilde{F}^* - dd^*||_{\ell_1(E)} \le \sigma, \qquad (12)$$
$$M \succeq 0.$$

The approximate fit is expressed in an entry-wise ℓ_1 sense. This feasibility problem is a convex program, for which there exist simple converging iterative methods. Once M is solved for, we have already seen that a model estimate can be obtained by extracting the leading eigenvector of M.

B. A practical algorithm

The convex formulation in (12) is too costly to solve at the scale of even toy problems. Let N be the total number of degrees of freedom of the unknown model m ; then the variable M of (12) is a $N \times N$ matrix, on which we want to impose positive semi-definiteness and approximate fit. To our knowledge, there is no time-efficient and memory-efficient algorithm to solve this type of semi-definite program when N ranges from 10^4 to 10^6.

We consider instead a non-convex relaxation of the feasibility problem (12), in which we limit the numerical rank of M to K, as in [6]. We may then write $M = RR^*$ where R is $N \times K$ and $K \ll N$. We replace the approximate ℓ_1 fit by Frobenius minimization. Regularization is also added to handle noise and uncertainty, yielding

$$\widehat{R} = \arg\min_R ||E \circ (\widetilde{F}RR^*\widetilde{F}^* - dd^*)||_F^2 + \lambda ||R||_F^2. \quad (13)$$

An estimate of m is obtained from $\widehat{R}$ by extracting the leading eigenvector of $\widehat{R}\widehat{R}^*$. Note that the Frobenius regularization on R is equivalent to a trace regularization on $M = RR^*$, which is known to promote the low-rank character of the matrix. The rank-K relaxation (13) can be seen as a generalization of the straightforward least-squares formulation (11). The two formulations coincide in the limit case $K = 1$. The strength of (13) is that the optimization variable is in a slightly bigger space than formulation (11).

The rank-K relaxation (13) is still nonconvex, but in practice, no local minimum has been observed even for $K = 2$, whereas the issue often arises for the least-squares approach (11). It should also be noted that the success of inversion is simple to test *a posteriori*, by comparing the sizes of the largest and second largest eigenvalues of M.

In practice, the memory requirement of rank-K relaxation is simply K times than that of non-lifted least-squares. The runtime per iteration is typically lower than K times that of a least-squares step however, because the bottleneck in the computation of $\widetilde{F}$ and $\widetilde{F}^*$ is usually the LU factorization of a Helmholtz operator, which needs to be done only once per iteration. Empirically, the number of iterations needed for convergence does not vary much as a function of K.

C. Examples

Our examples fall into three categories.

1) **Linear inverse source problem.** Here we consider a set of receiver locations x_r, and a constant density acoustics inverse source problem, which reads in the Fourier domain

$$-(\Delta + \omega^2 m_0(x))\widehat{u}_s(x,\omega) = \widehat{w}(\omega)m(x)$$

$$Fm = d(x_r,\omega) = \widehat{u}_s(x_r,\omega)$$

$$m_0(x) = \frac{1}{c_0(x)^2} \qquad \text{(squared slowness)}$$

Waves are propagating from a source term with known time signature w. The problem is to reconstruct the spatial distribution m.

2) **Linearized inverse scattering problem.** Here we consider a set of receivers x_r, waves generated by sources x_s, and a constant density acoustic inverse problem for the reflectivity perturbation m_1,

$$-(\Delta + \omega^2 m_0(x))\widehat{u}_{0,s}(x,\omega) = \widehat{w}(\omega)\delta(x - x_s)$$
$$-(\Delta + \omega^2 m_0(x))\widehat{u}_{1,s}(x,\omega) = \omega^2\widehat{u}_{0,s}(x,\omega)m_1(x)$$

$$Fm_1 = d(x_r,x_s,\omega) = \widehat{u}_{1,s}(x_r,\omega).$$

The isolation of the Born scattered wavefield (primary reflections) $u_{1,s}$ from the full scattered field, although a very difficult task in practice, is assumed to be performed perfectly in this paper.

3) **Full waveform inversion (FWI).** We again consider a set of receivers x_r, waves generated by sources x_s, and a constant density acoustic inverse problem for the reflectivity m, with

$$-(\Delta + \omega^2 m(x))\widehat{u}_s(x,\omega) = \widehat{w}(\omega)\delta(x - x_s)$$

$$\mathcal{F}(m) = d(x_r,x_s,\omega) = \widehat{u}_s(x_r,\omega).$$

In a first numerical experiment (Figure 4), we consider a linearized inverse scattering problem where $c = 1/\sqrt{m}$ is the Marmousi2 p-wave velocity model [22], m_0 is a smoothed version of m, and $m_1 = m - m_0$ is the model perturbation used to generated data in the linearized forward model. The sources and receivers are placed at the top of the image in an equispaced fashion, with 30 sources and 300 receivers. The frequency sampling is uniform between 3 and 10 Hz, with 10 frequency samples. The Helmholtz equation is discretized with fourth-order finite differences at about 20 points per wavelength for data modeling, while the simulations for the inversion are with second-order finite differences. The noise model is taken to be Gaussian,

$$\widetilde{d}_i = d_i + \eta_i \quad \eta_i \sim \mathbb{CN}(0,\sigma^2)$$

where $\sigma = 0.1\frac{||b||_2}{\sqrt{2n}}$, so that $\frac{||\eta||_2}{||b||_2} = 0.1$ (10% additive noise). Figure 4 shows stable recovery of m_1 from noisy d, both by least-squares and by interferometric inversion. In this case the graph E is taken to be an Erdős-Rényi random graph with $p = 1.5\frac{log(N)}{N}$ to ensure connectedness. (Instabilities occur if p is smaller; larger values of p do not substantially help.) Note that if E were chosen as a disconnected graph that only forms cross-correlations (same ω for the data indices i and j), then interferometric inversion is not well-posed and does not result in good images (not shown). The optimization method for interferometric inversion is the rank-2 relaxation scheme mentioned earlier. The message of this numerical example is twofold: it shows that stable recovery is possible in the interferometric regime, when E is properly chosen; and a comparison of Figures 4 bottom and middle shows that it does not come with the loss of resolution that would be expected from CINT [4]. This observation was also a key conclusion in the recent paper by Mason et al. [23].

In a second numerical experiment (Figure 5), we consider full waveform inversion where $c = 1/\sqrt{m}$ is the Marmousi2 p-wave velocity model, and the initial model is the same m_0 as used in the previous example. The setup for the acquisition, the Helmholtz equation, the noise, and the mask E are the same as before. Figure 5 shows the result of quasi-Newton (LBFGS) iterations

with a frequency sweep, and the corresponding interfero-
metric inversion result. To produce this latter result, we
simply replace the classical adjoint-state gradient step
by interferometric inversion applied to the data residual.
In all the numerical experiments so far (Figures 4 and
5), the values of the model misfits are not meaningfully
different in the least-squares and interferometric cases.
The message of this numerical example is that there
seems to be no loss of resolution in the full waveform
case either, when comparing Figure 5 (bottom) to its
least-squares counterpart (second from bottom).

Possible limitations of the interferometric approach
are the slightly higher computational and memory cost
(as discussed in the previous section); the need to prop-
erly choose the data mask E; and the fact that the data
match is now quartic rather than quadratic in the original
data d. Quartic objective functions can be problematic in
the presence of outliers, such as when the noise is heavy-
tailed, because they increase the relative importance of
those corrupted measurements. We did not attempt to
deal with heavy-tailed noise in this paper.

It is also worth noting that "backprojecting the cross-
correlations" (or the more general quadratic combina-
tions we consider here) can be shown to be related to the
first iteration of gradient descent in the *lifted semidefinite
formulation* of interferometric inversion.

So far, our numerical experiments merely confirm the
theoretical prediction that interferometric inversion can
be accurate and stable under minimal assumptions on the
mask/graph E of data pair products. In the next section,
we show that there is also an important *rationale* for
switching to the interferometric formulation: its results
display robustness vis-a-vis some uncertainties in the
forward model F.

D. Model robustness

In this section we continue to consider wave-based
imaging scenarios, but more constrained in the sense
that the receivers and/or sources completely surround
the object to be imaged, i.e., provide a full aperture.
The robustness claims below are only for this case. On
the other hand, we now relax the requirement that the
forward model $\widetilde{F}$ used for inversion be identical or very
close to the forward map F used for data modeling.
The interferometric mask E is also different, and more
physically meaningful than in the previous section: it
indicates a small band around the diagonal $i = j$, i.e., it
selects nearby receiver positions and frequencies, as in
[4]. The parameters of this selector were tuned for best
results; they physically correspond to the idea that two
data points d_i and d_j should only be compared if they are
within one cycle of one another for the most oscillatory
phase in d. In all our experiments with model errors,
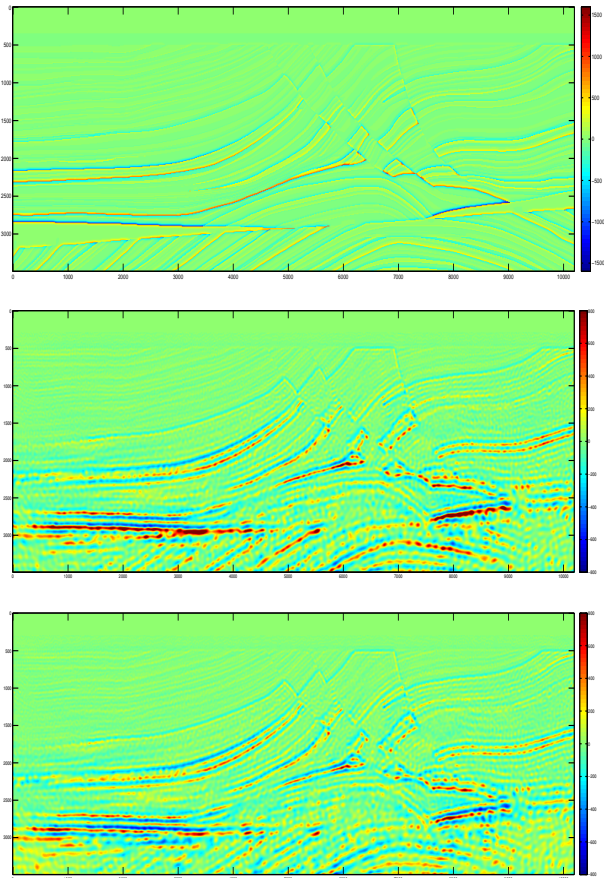it is crucial that the data misfit parameter σ be positive

Fig. 4. Top: true, unknown reflectivity profile m_1 used to generate
data. Middle: least-squares solution. Bottom: result of interferometric
inversion. This example shows that interferometric inversion is stable
when the mask E is connected, as in the theory, and shows no apparent
loss of resolution vs. least squares.

nonzero – it would be a mistake to try and explain the
data perfectly with a wrong model.

In the next numerical experiment (Figure 6), we con-
sider the inverse source problem, where the source distri-
bution m is the Shepp-Logan phantom. This community
model is of interest in ultrasound medical imaging,
where it represents a horizontal cross-section of different
organs in the torso. The receivers densely surround the
phantom and are depicted as white crosses. Equispaced
frequencies are considered on the bandwidth of w. A
small modeling error is assumed to have been made on
the background velocity: in the experiment, the waves
propagated with unit speed $c_0(x) = 1$, but in the simu-
lation, the waves propagate more slowly, $\widetilde{c}_0(x) = 0.95$.
As shown in Figure 6 (middle), least-squares inversion
does not properly handle this type of uncertainty and
produces a defocused image. (Least-squares inversion
is regularized with the ℓ_2 norm of the model, a.k.a.
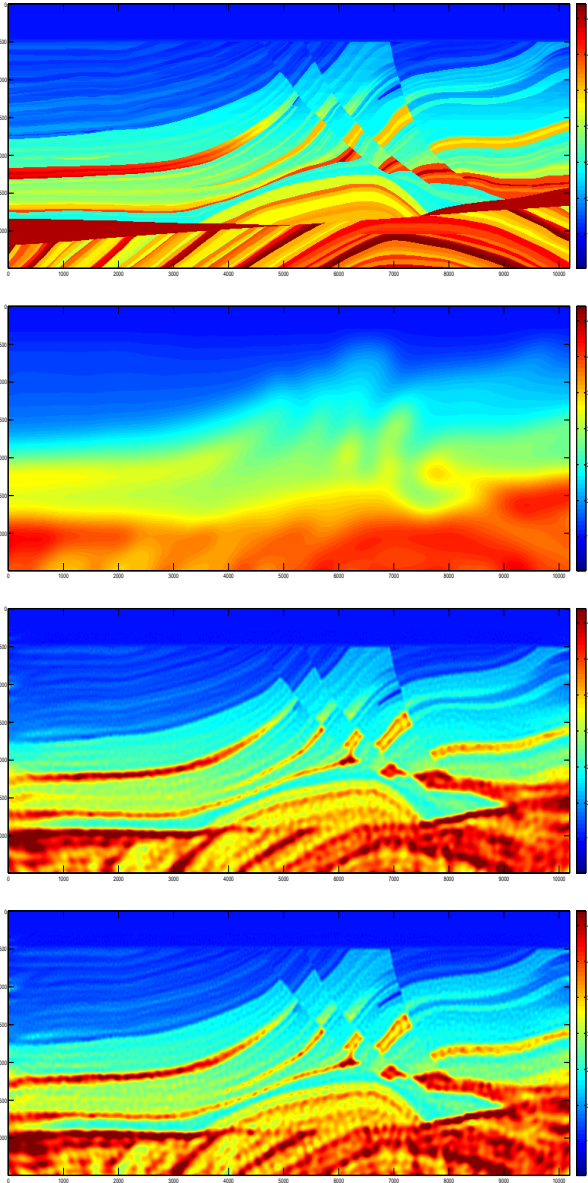Tykhonov. A large range of values of the regularization

Fig. 5. Top: true, unknown map of the squared slowness m used to generate data. Second: initial guess for either inversion scheme. Third: nonlinear least-squares solution, a.k.a. full waveform inversion. Bottom: result of interferometric inversion. This example shows that the stability and resolution properties of interferometric inversion carry over to the case of full waveform inversion.

parameter was tested, and the best empirical results are reported here.) In contrast, interferometric inversion, shown in Figure 6 (bottom), enjoys a better resolution. The price to pay for focusing is positioning: although we do not have a mathematical proof of this fact, the interferometric reconstruction is near a shrunk version of the true source distribution.

In the last numerical experiment (Figure 7), we consider the linearized inverse scattering problem, where

the phantom is now the reflectivity perturbation m_1. As the figure shows, sources and receivers surround the phantom, with a denser sampling for the receivers than for the sources. The wave speed profile is uniform ($c = 1$). In this example, the modeling error is assumed to be on the receiver positions: they have a smooth random deviation from a circle, as shown in Figure 7 (top). Again, least-squares inversion produces a poor result regardless of the Tykhonov regularization parameter (Figure 7, middle), where the features of the phantom are not clearly reconstructed, and the strong outer layer is affected by a significant oscillatory error. Interferometric inversion produces a somewhat more focused reconstruction (Figure 7, bottom), where more features of the phantom are recovered and the outer layer is well-resolved.

Model robustness is heuristically plausible in the scenario when data are of the form $d_i \sim e^{i\omega\tau_i}$ for some traveltimes τ_i which are themselves function of a velocity c through $\tau_i = \delta_i/c$. In that case, the combination $d_i\overline{d_j} \sim e^{i\omega(\tau_i-\tau_j)}$ has a phase that encodes the idea of a *traveltime difference*. When i is near j in data space (because they correspond, say, to nearby receivers), it is clear that $\tau_i - \tau_j$ depends in a milder fashion on errors in c, or on correlated errors in δ_i, than the individual traveltimes themselves. This results in some degree of stability of selected $d_i\overline{d_j}$ with respect to those types of errors, which in turn results in more focused imaging. This phenomenon is not unlike the celebrated statistical stability of $d_i\overline{d_j}$ under some models of randomness on either the velocity or the sources. For random media, this behavior was leveraged in the context of coherent interferometric imaging (CINT) in the 2011 work of Borcea et al. [5]

IV. DISCUSSION

This paper explores a recent optimization idea, convex relaxation by lifting, for dealing with the quadratic data combinations that occur in the context of interferometry. The role of the mask E that selects the quadratic measurements is explained: it is shown that it should encode a well-connected graph for robustness of the recovery. The recovery theorems do not assume a random model on this graph; instead, they involve the Laplacian's spectral gap.

Numerical experiments in the context of imaging are shown to confirm that recovery is stable when the assumptions of the theorem are obeyed. Developing methods to solve the lifted problem at interesting scales is a difficult problem that we circumvent via an ad-hoc rank-2 relaxation scheme. It is observed empirically that there is no noticeable loss of resolution from switching to an interferometric optimization formulation. It is also observed that interferometric inversion for imaging can

display a puzzling degree of robustness to very specific model uncertainties.

We leave numerous important questions unanswered, such as how to optimize the choice of the selector E when there is a trade-off between connectivity and robustness to model uncertainties; how to select the data misfit parameter $\sigma > 0$ and whether there is a phase transition in σ for model robustness vs. lack thereof; how to justify the gain in focusing in some reconstructions; and whether this gain in focusing is part of a tradeoff with the geometric faithfulness of the recovered image.

APPENDIX

A. Proof of theorem 1.

Observe that if x is feasible for (3), then xx^* is feasible for (4), and has $e^{i\alpha}x$ as leading eigenvector. Hence we focus without loss of generality on (4).

As in [30], consider the Laplacian matrix weighted with the unknown phases,

$$\mathcal{L} = \Lambda L \Lambda^*,$$

with $\Lambda = \mathrm{diag}(x_0)$. In other words $\mathcal{L}_{ij} = (X_0)_{ij} L_{ij}$ with $X_0 = x_0 x_0^*$. We still have $\mathcal{L} \succeq 0$ and $\lambda_1 = 0$, but now $v_1 = \frac{1}{\sqrt{n}} x_0$. Here and below, λ and v refer to $\mathcal{L}$, and v has unit ℓ_2 norm.

The idea of the proof is to compare X with the rank-1 spectral projectors $v_j v_j^*$ of $\mathcal{L}$. Let $\langle A, B \rangle = \mathrm{tr}(AB^*)$ be the Frobenius inner product. Any X obeying (3) can be written as $X = X_0 + \widetilde{\varepsilon}$ with $\|\widetilde{\varepsilon}\|_1 \leq \|\varepsilon\|_1 + \sigma$. We have

$$\langle X, \mathcal{L} \rangle = \langle X_0, \mathcal{L} \rangle + \langle \widetilde{\varepsilon}, \mathcal{L} \rangle$$

A short computation shows that

$$\langle X_0, \mathcal{L} \rangle = \sum_i (X_0)_{ii} \overline{\mathcal{L}_{ii}} + \sum_{(i,j) \in E} (X_0)_{ij} \overline{\mathcal{L}_{ij}}$$
$$= -\sum_i d_i + \sum_{(i,j) \in E} (x_0)_i \overline{(x_0)_j} \, \overline{(x_0)_i} (x_0)_j$$
$$= \sum_i \left[-d_i + \sum_{j:(i,j) \in E} 1 \right]$$
$$= 0.$$

Since $|L_{ij}| = 1$ on E, the error term is simply bounded as

$$|\langle \widetilde{\varepsilon}, \mathcal{L} \rangle| \leq \|\widetilde{\varepsilon}\|_1$$

On the other hand the Laplacian expands as

$$\mathcal{L} = \sum_j v_j \lambda_j v_j^*,$$

so we can introduce a convenient normalization factor $1/n$ and write

$$\left\langle \frac{X}{n}, \mathcal{L} \right\rangle = \sum_j c_j \lambda_j, \tag{14}$$

with

$$c_j = \left\langle \frac{X}{n}, v_j v_j^* \right\rangle = \frac{v_j^* X v_j}{n}.$$

Notice that $c_j \geq 0$ since we require $X \succeq 0$. Their sum is

$$\sum_j c_j = \left\langle \frac{X}{n}, \sum_j v_j v_j^* \right\rangle = \left\langle \frac{X}{n}, I \right\rangle = \frac{\mathrm{tr}(X)}{n} = 1.$$

Hence (14) is a convex combination of the eigenvalues of $\mathcal{L}$, bounded by $\|\widetilde{\varepsilon}\|_1/n$. The smaller this bound, the more lopsided the convex combination toward λ_1, i.e., the larger c_1. The following lemma makes this observation precise.

Lemma 1. *Let $\mu = \sum_j c_j \lambda_j$ with $c_j \geq 0$, $\sum_j c_j = 1$, and $\lambda_1 = 0$. If $\mu \leq \lambda_2$, then*

$$c_1 \geq 1 - \frac{\mu}{\lambda_2}.$$

Proof of lemma 1..

$$\mu = \sum_{i \geq 2} c_j \lambda_j \geq \lambda_2 \sum_{j \geq 2} c_j = \lambda_2 (1 - c_1),$$

then isolate c_1. $\qquad\square$

Assuming $\|\widetilde{\varepsilon}\|_1 \leq n\lambda_2$, we now have

$$\left\langle \frac{X}{n}, v_1 v_1^* \right\rangle \geq 1 - \frac{\|\widetilde{\varepsilon}\|_1}{n\lambda_2}.$$

We can further bound

$$\left\| \frac{X}{n} - v_1 v_1^* \right\|_F^2 = \mathrm{tr}\left[\left(\frac{X}{n} - v_1 v_1^* \right)^2 \right]$$
$$= \mathrm{tr}((v_1 v_1^*)^2) + \frac{\mathrm{tr}(X^2)}{n^2} - 2\,\mathrm{tr}\left(\frac{X}{n} v_1 v_1^* \right).$$

The first term is 1. The second term is less than 1, since $\mathrm{tr}(X^2) \leq \mathrm{tr}(X)^2$ for positive semidefinite matrices. Therefore,

$$\left\| \frac{X}{n} - v_1 v_1^* \right\|_F^2 \leq 2 - 2\,\mathrm{tr}\left(\frac{X}{n} v_1 v_1^* \right) \leq \frac{2\|\widetilde{\varepsilon}\|_1}{n\lambda_2}.$$

We can now control the departure of the top eigenvector of X/n from v_1 by the following lemma. It is analogous to the sin theta theorem of Davis-Kahan, except for the choice of normalization of the vectors. (It is also a generalization of a lemma used by one of us in [11] (section 4.2).) The proof is only given for completeness.

Lemma 2. *Consider any Hermitian $X \in \mathbb{C}^{n \times n}$, and any $v \in \mathbb{C}^n$, such that $\|X - vv^*\| < \frac{\|v\|^2}{2}$. Let η_1 be the leading eigenvalue of X, and x_1 the corresponding unit-norm eigenvector. Let x be defined either as (a) $x_1\|v\|$, or as (b) $x_1\sqrt{\eta_1}$. Then*

$$\| \, x\|x\| - e^{i\alpha} v\|v\| \, \| \leq 2\sqrt{2} \, \|X - vv^*\|,$$

for some $\alpha \in [0, 2\pi)$.

Proof of Lemma 2. Let $\delta = \|X - vv^*\|$. Notice that $\|vv^*\| = \|v\|^2$. Decompose $X = \sum_{j=1}^{n} x_j \eta_j x_j^*$ with eigenvalues η_j sorted in decreasing order. By perturbation theory for symmetric matrices (Weyl's inequalities),

$$\max\{|\|v\|^2 - \eta_1|, |\eta_2|, \ldots, |\eta_n|\} \le \delta, \qquad (15)$$

so it is clear that $\eta_1 > 0$, and that the eigenspace of η_1 is one-dimensional, as soon as $\delta < \frac{\|v\|^2}{2}$.

Let us deal first with the case (a) when $x = x_1 \|v\|$. Consider

$$vv^* - xx^* = vv^* - X + Y,$$

where

$$Y = x_1(\|v\|^2 - \eta_1)x_1^* + \sum_{j=2}^{n} x_j \eta_j x_j^*.$$

From (15), it is clear that $\|Y\| \le \delta$. Let $v_1 = v/\|v\|$. We get

$$\|vv^* - xx^*\| \le \|vv^* - X\| + \|Y\| \le 2\delta.$$

Pick α so that $|v^* x| = e^{-i\alpha} v^* x$. Then

$$
\begin{aligned}
&\| v\|v\| - e^{-i\alpha} x\|x\| \|^2 \\
&= \|v\|^4 + \|x\|^4 - 2\|v\|\,\|x\|\,\Re\, e^{-i\alpha} v^* x \\
&= \|v\|^4 + \|x\|^4 - 2\|v\|\,\|x\|\,|v^* x| \qquad \text{by def. of } \alpha \\
&\le \|v\|^4 + \|x\|^4 - 2\,|v^* x|^2 \qquad \text{by Cauchy-Schwarz} \\
&= \|vv^* - xx^*\|_F^2 \\
&\le 2\|vv^* - xx^*\|^2 \qquad \text{since } vv^* - xx^* \text{ has rank 2} \\
&\le 8\delta^2.
\end{aligned}
$$

The case (b) when $x = x_1\sqrt{\eta_1}$ is treated analogously. The only difference is now that

$$Y = \sum_{j=2}^{n} x_j \eta_j x_j^*.$$

A fortiori, $\|Y\| \le \delta$ as well. $\qquad\square$

Part (a) of lemma 2 is applied with X/n in place of X, and v_1 in place of v. In that case, $\|v_1\| = 1$. We conclude the proof by noticing that $v_1 = \frac{x_0}{\sqrt{n}}$, and that the output x of the lifting method is normalized so that $x_1 = \frac{x}{\sqrt{n}}$.

B. Proof of theorem 3

The proof is a simple argument of perturbation of eigenvectors. We either assume $\varepsilon_{ij} = \overline{\varepsilon_{ji}}$ or enforce it by symmetrizing the measurements. Define $\mathcal{L}$ as previously, and notice that $\|\mathcal{L} - \widetilde{\mathcal{L}}\| \le \|\varepsilon\|$. Consider the eigen-decompositions

$$\mathcal{L} v_j = \lambda_j v_j, \qquad \widetilde{\mathcal{L}} \widetilde{v}_j = \widetilde{\lambda}_j \widetilde{v}_j,$$

with $\lambda_1 = 0$. Form

$$\widetilde{\mathcal{L}} v_j = \lambda_j v_j + r_j,$$

with $\|r_j\| \le \|\varepsilon\|$. Take the dot product of the equation above with $\widetilde{v}_k$ to obtain

$$\langle \widetilde{v}_k, r_j \rangle = (\widetilde{\lambda}_k - \lambda_j)\langle \widetilde{v}_k, v_j \rangle.$$

Let $j = 1$, and use $\lambda_1 = 0$. We get

$$\sum_{k \ge 2} |\langle \widetilde{v}_k, v_1 \rangle|^2 \le \frac{\sum_{k \ge 2} |\langle \widetilde{v}_k, r_1 \rangle|^2}{\max_{k \ge 2} |\widetilde{\lambda}_k|^2} \le \frac{\|\varepsilon\|^2}{\widetilde{\lambda}_2^2}.$$

As a result,

$$|\langle \widetilde{v}_1, v_1 \rangle|^2 \ge 1 - \frac{\|\varepsilon\|^2}{\widetilde{\lambda}_2^2}.$$

Choose α so that $\langle e^{i\alpha}\widetilde{v}_1, v_1 \rangle = |\langle \widetilde{v}_1, v_1 \rangle|$. Then

$$
\begin{aligned}
\|v_1 - e^{i\alpha}\widetilde{v}_1\|^2 &= 2 - 2\Re\langle e^{i\alpha}\widetilde{v}_1, v_1 \rangle \\
&= 2 - 2|\langle \widetilde{v}_1, v_1 \rangle| \\
&\le 2 - 2|\langle \widetilde{v}_1, v_1 \rangle|^2 \\
&\le 2\frac{\|\varepsilon\|^2}{\widetilde{\lambda}_2^2}.
\end{aligned}
$$

Conclude by multiplying through by n and taking a square root.

C. Proof of theorem 4.

The proof follows the argument in section A; we mostly highlight the modifications.

Let $b_i = |b_i| e^{i\phi_i}$. The Laplacian with phases is $\mathcal{L}_b = \Lambda_\phi L_{|b|} \Lambda_\phi^*$, with $\Lambda_\phi = \text{diag}(e^{i\phi_i})$. Explicitly,

$$
(\mathcal{L}_b)_{ij} = \begin{cases}
\sum_{k:(i,k)\in E} |b_k|^2 & \text{if } i = j; \\
-b_i \overline{b_j} & \text{if } (i,j) \in E; \\
0 & \text{otherwise,}
\end{cases}
$$

The matrix $Y = AXA^*$ is compared to the rank-1 spectral projectors of $\mathcal{L}_b$. We can write it as $Y = bb^* + \widetilde{\varepsilon}$ with $\|\widetilde{\varepsilon}\|_1 \le \|\varepsilon\|_1 + \sigma$. The computation of $\langle bb^*, \mathcal{L}_b \rangle$ is now

$$
\begin{aligned}
\langle bb^*, \mathcal{L}_b \rangle &= \sum_i b_i \overline{b_i} \overline{\mathcal{L}_{ii}} + \sum_{(i,j)\in E} b_i \overline{b_j}\, \overline{\mathcal{L}_{ij}} \\
&= -\sum_i |b_i|^2 \sum_{k:(i,k)\in E} |b_k|^2 + \sum_{(i,j)\in E} b_i \overline{b_j}\, \overline{b_i} b_j \\
&= \sum_i |b_i|^2 \left[-\sum_{j:(i,j)\in E} |b_j|^2 + \sum_{j:(i,j)\in E} |b_j|^2 \right] \\
&= 0.
\end{aligned}
$$

The error term is now bounded (in a rather crude fashion) as

$$|\langle \widetilde{\varepsilon}, \mathcal{L}_b \rangle| \leq \|\mathcal{L}_b\|_\infty \|\widetilde{\varepsilon}\|_1$$

$$\leq \left[\max_i \sum_{j:(i,j)\in E} |b_j|^2 \right] \|\widetilde{\varepsilon}\|_1 \leq \|b\|^2 \|\widetilde{\varepsilon}\|_1.$$

Upon normalizing Y to unit trace, we get

$$|\langle \frac{Y}{\operatorname{tr}(Y)}, \mathcal{L}_b \rangle| \leq \frac{\|b\|^2 \|\widetilde{\varepsilon}\|_1}{\|b\|^2 + \operatorname{tr}(\widetilde{\varepsilon})} \leq 2\|\widetilde{\varepsilon}\|_1,$$

where the last inequality follows from

$$|\operatorname{tr}(\widetilde{\varepsilon})| \leq \|\widetilde{\varepsilon}\|_1$$
$$\leq \|\varepsilon\|_1 + \sigma$$
$$\leq \lambda_2/2 \quad \text{(assumption of the theorem)}$$
$$\leq \|b\|^2/2 \quad \text{(by Gershgorin).}$$

On the other hand, we expand

$$\langle \frac{Y}{\operatorname{tr}(Y)}, \mathcal{L}_b \rangle = \sum_j c_j \lambda_j,$$

and use $X \succeq 0 \Rightarrow Y \succeq 0$ to get $c_j \geq 0$, $\sum_j c_j = 1$. Since $2\|\widetilde{\varepsilon}\|_1 \leq \lambda_2$, we conclude as in section A that

$$\langle \frac{Y}{\operatorname{tr}(Y)}, v_1 v_1^* \rangle \geq 1 - \frac{2\|\widetilde{\varepsilon}\|_1}{\lambda_2},$$

hence

$$\|\frac{Y}{\operatorname{tr}(Y)} - v_1 v_1^*\|_F^2 \leq 4\frac{\|\widetilde{\varepsilon}\|_1}{\lambda_2}. \tag{16}$$

For $X = A^+ Y (A^*)^+$, we get

$$\|\frac{X}{\operatorname{tr}(Y)} - (A^+ v_1)(A^+ v_1)^*\|_F^2 \leq 4\|A^+\|^4 \frac{\|\widetilde{\varepsilon}\|_1}{\lambda_2}.$$

Call the right-hand side δ^2. Recall that $v_1 = b/\|b\|$ hence $A^+ v_1 = x_0/\|b\|$. Using $\operatorname{tr}(Y) = \|b\|^2 + \operatorname{tr}(\widetilde{\varepsilon})$, we get

$$\|X - x_0 x_0^*\| \leq \delta \operatorname{tr}(Y) + \frac{\|x_0\|^2}{\|b\|^2} |\operatorname{tr}(\widetilde{\varepsilon})|. \tag{17}$$

Elementary calculations based on the bound $\|\widetilde{\varepsilon}\|_1 \leq \lambda_2/2 \leq \|b\|^2/2$ allow to further bound the above quantity by $\frac{(6+\sqrt{2})}{4}\delta\|b\|^2$. We can now call upon lemma 2, part (b), to obtain

$$\|x\|x\| - e^{i\alpha} x_0 \|x_0\|\| \leq 2\sqrt{2}\frac{(6+\sqrt{2})}{4}\delta\|b\|^2,$$

where $x = x_1 \sqrt{\lambda_1(X)}$ is the leading eigenvector of X normalized so that $\|x\|^2 = \lambda_1(X)$ is the leading eigenvalue of X. We use (17) one more time to bound

$$|\lambda_1(X) - \|x_0\|^2| \leq \frac{(6+\sqrt{2})}{4}\delta\|b\|^2,$$

hence

$$\|x_0\| \|x - e^{i\alpha} x_0\|$$
$$\leq \|x\|x\| - e^{i\alpha} x_0\|x_0\| + \|x\| \|\|x\| - \|x_0\|\|$$
$$\leq 2\sqrt{2}\frac{(6+\sqrt{2})}{4}\delta\|b\|^2 + \frac{\|x\|}{\|x\| + \|x_0\|}\|\|x\|^2 - \|x_0\|^2\|$$
$$\leq (2\sqrt{2}+1)\frac{(6+\sqrt{2})}{4}\delta\|b\|^2.$$

Use $\|b\| \leq \|A\| \|x_0\|$ and the formula for δ to conclude that

$$\|x - e^{i\alpha} x_0\| \leq C \|x_0\| \kappa(A)^2 \sqrt{\frac{\|\widetilde{\varepsilon}\|_1}{\lambda_2}},$$

with $C = 2(2\sqrt{2}+1)\frac{(6+\sqrt{2})}{4} \leq 15$.

D. Proof of theorem 5.

The proof proceeds as in the previous section, up to equation (16). The rest of the reasoning is a close mirror of the one in the previous section, with Y in place of X, y in place of x, b in place of x_0, and δ re-set to $2\sqrt{\|\widetilde{\varepsilon}\|_1/\lambda_2}$. We obtain

$$\|y - e^{i\alpha} b\| \leq 15 \|b\| \sqrt{\frac{\|\widetilde{\varepsilon}\|_1}{\lambda_2}}.$$

We conclude by letting $x = A^+ y$, $x_0 = A^+ b$, and using $\|b\| \leq \|A\|\|x_0\|$.

REFERENCES

[1] B. Alexeev, A. S. Bandeira, M. Fickus and D. G. Mixon. Phase retrieval with polarization, *SIAM J. Imaging Sci.*, 7(1), 3566
[2] A. S. Bandeira, A. Singer, and D. A. Spielman. A Cheeger inequality for the graph connection Laplacian, *SIAM. J. Matrix Anal. Appl.*, 34(4), 16111630
[3] P. Blomgren, G. Papanicolaou, and H. Zhao, Super-resolution in time-reversal acoustics, *J. Acoust. Soc. Am.* 111(1), 230-248, 2002
[4] L. Borcea, G. Papanicolaou, and C. Tsogka, Interferometric array imaging in clutter, *Inv. Prob.* 21(4), 1419-1460, 2005
[5] L. Borcea, J. Garnier, G. Papanicolaou and C. Tsogka, Enhanced Statistical Stability in Coherent Interferometric Imaging, *Inverse Problems* 27 (2011) 085004.
[6] Burer, S., and R. Monteiro, 2003, A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization: Mathematical Programming (Series B), **95**, 329–357.
[7] E. J. Candes, Y. C. Eldar, T. Strohmer and V. Voroninski. Phase retrieval via matrix completion, *SIAM J. Imaging Sci.*, 6(1), 199-225, 2013
[8] E. J. Candes, B. Recht, Exact Matrix Completion via Convex Optimization, *Found. Comput. Math.*, 9-6, 717-772, 2009
[9] A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements, *Inverse Problems 27.1 015005*, 2011
[10] S. Chatterjee, Matrix estimation by Universal Singular Value Thresholding, *Ann. Statist.* Volume 43, Number 1 (2015), 177-214
[11] L. Demanet and P. Hand. Stable optimizationless recovery from phaseless linear measurements, *Journal of Fourier Analysis and Applications*, Volume 20, Issue 1, (2014) pp 199221

[12] E. Dussaud, *Velocity analysis in the presence of uncertainty*, Ph.D. thesis, Computational and Applied Mathematics, Rice University, 2005

[13] A. Fichtner, Source and Processing Effects on Noise Correlations, *Geophysical J. Int.* 197 (3): 1527?31, 2014

[14] J. Garnier, Imaging in randomly layered media by cross-correlating noisy signals, *SIAM Multiscale Model. Simul.* 4, 610-640, 2005

[15] M. Goemans and D. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *Journal of the ACM*, 42(6),1115-1145, 1995

[16] S. Hanasoge, Measurements and kernels for source-structure inversions in noise tomography, *Geophysical J. Int.* 196 (2): 971- 985, 2014

[17] A. Javanmard, A. Montanari, Localization from incomplete noisy distance measurements, *Found. Comput. Math.* 13, 297-345, 2013

[18] V. Jugnon and L. Demanet, Interferometric inversion: a robust approach to linear inverse problems, in *Proc. SEG Annual Meeting, 2013*

[19] R. Keshavan, A. Montanari, S. Oh, Matrix Completion from Noisy Entries, *Journal of Machine Learning Research* 11, 2057-2078, 2010

[20] O. I. Lobkis and R. L. Weaver, On the emergence of the Greens function in the correlations of a diffuse field, *J. Acoustic. Soc. Am.*, 110, 3011-3017, 2001

[21] L. Lovasz, *Eigenvalues of graphs*, Lecture notes, 2007

[22] G. Martin, R. Wiley, K. Marfurt, Marmousi2: An elastic upgrade for Marmousi, *The Leading Edge, Society for Exploration Geophysics*, **25**, 2006, 156-166

[23] E. Mason, I. Son, B. Yazici, Passive Synthetic Aperture Radar Imaging Using Low-Rank Matrix Recovery Methods, *IEEE Journal of Selected Topics in Signal Processing*, 9-8, 2015

[24] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Review* 52-3, 471-501, 2010

[25] J. M. Rodenburg, A. C. Hurst, A. G. Cullis, B. R. Dobson, F. Pfeiffer, O. Bunk, C. David, K. Jefimovs, and I. Johnson. Hard-x-ray lensless imaging of extended objects, *Physical review letters 98, no. 3 034801*, 2007

[26] Sava, P., and O. Poliannikov, 2008, Interferometric imaging condition for wave-equation migration: Geophysics, **73**, S47–S61.

[27] J. Schotland, Quantum imaging and inverse scattering, *Optics letters*, 35(20), 3309-3311, 2010

[28] G. T. Schuster, *Seismic Interferometry* Cambridge University Press, 2009

[29] G. T. Schuster, J. Yu, J. Sheng, and J. Rickett, Interferometric/daylight seismic imaging *Geophysics* 157(2), 838-852, 2004

[30] A. Singer, Angular synchronization by eigenvectors and semidefinite programming, *Applied and computational harmonic analysis 30.1 20-36*, 2011

[31] A. Singer and M. Cucuringu, Uniqueness of low-rank matrix completion by rigidity theory, *SIAM. J. Matrix Anal. Appl.* 31(4), 1621-1641, 2010

[32] J. Tromp, Y. Luo, S. Hanasoge, and D. Peter, Noise cross-correlation sensitivity kernels, *Geophysical J. Int.* 183, 791? 819, 2010

[33] L. Wang and A. Singer. Exact and Stable Recovery of Rotations for Robust Synchronization, *Information and Inference: A Journal of the IMA* (2013) 2, 145193

[34] I. Waldspurger, A. d'Aspremont, and S. Mallat, Phase recovery, maxcut and complex semidefinite programming, *Mathematical Programming*, Volume 149, Issue 1 (2015), pp 4781

[35] K. Wapenaar and J. Fokkema, Greens function representations for seismic interferometry, *Geophysics*, 71, SI33-SI46, 2006
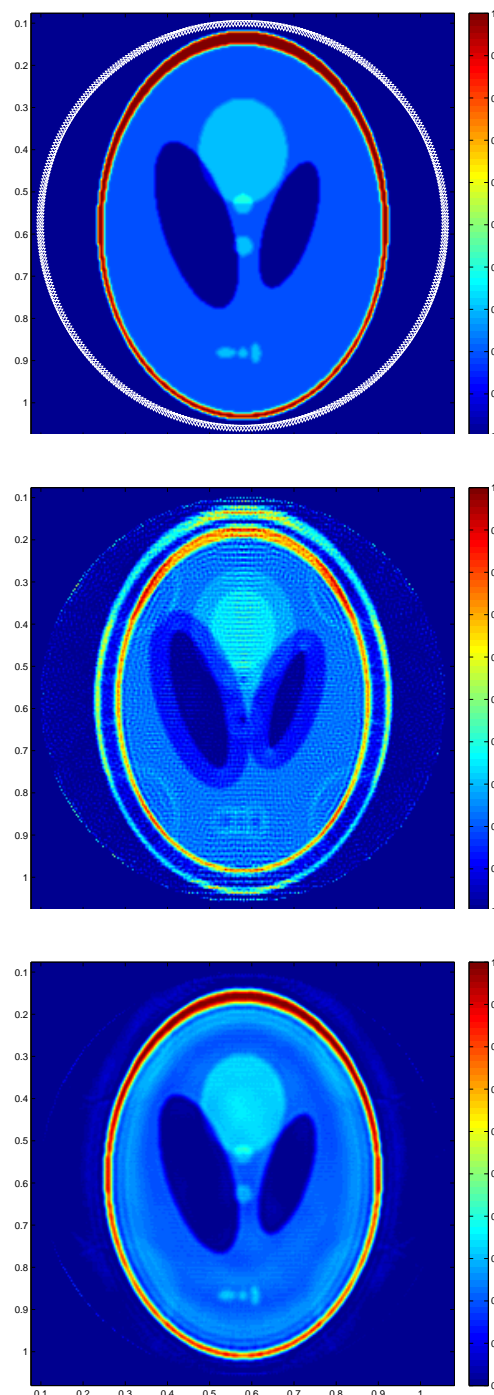
Fig. 6. Top: Setting of the inverse source experiment. Middle: least-squares reconstruction. Bottom: interferometric reconstruction. This example shows model robustness for focused imaging, in the specific case of full aperture, when the underlying constant wave speed is slightly wrong in the forward model.
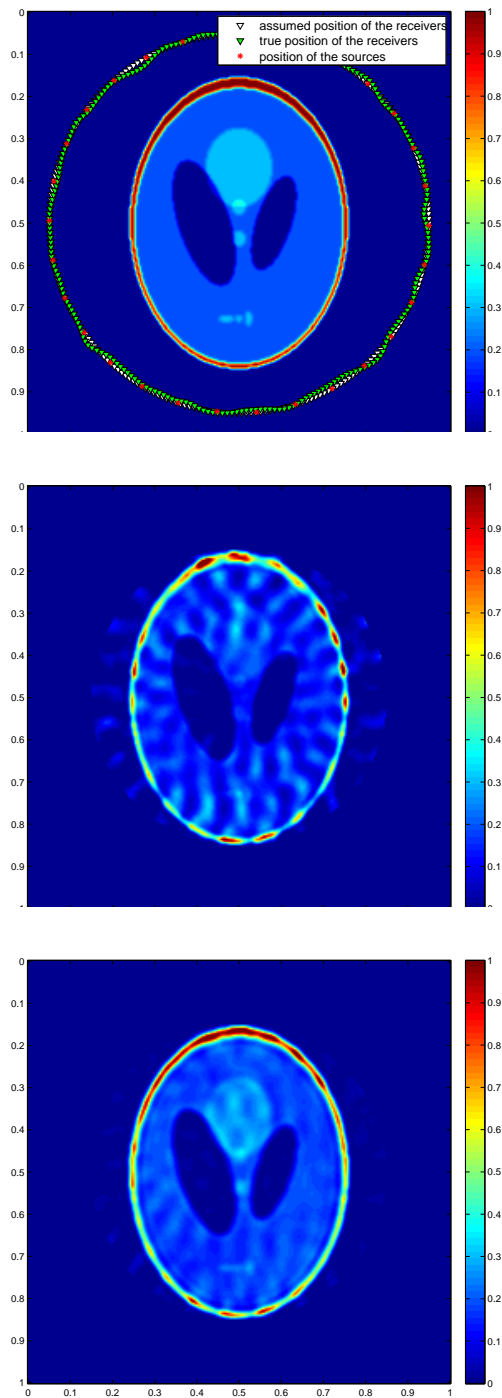
Fig. 7. Top: Setting of the inverse scattering experiment. Middle: least-squares reconstruction. Bottom: interferometric reconstruction. This example shows model robustness for focused imaging, in the specific case of full aperture, when the receiver locations are slightly wrong in the forward model.