TOWARDS OPTIMAL SENSOR PLACEMENT FOR SPARSE INVERSE PROBLEMS WITH RANDOM NOISE

DANIEL WALTER

Looking at the lifework of Richard von Mises, one might be baffled by the diverse range of topics that he left his footprint on. Among others, he contributed greatly to aero- and hydrodynamics, plasticity, probability theory and statistics but also the philosophy of sciences. Much of his mathematical contribution was influenced by his mechanical works; always striving to connect applied mathematics to real-world problems. In this talk, we want to celebrate the diversity of von Mises' work by addressing mathematically a problem from engineering science using a wide array of tools from different mathematical fields such as nonsmooth analysis, variational regularization theory and statistical inference, [3].

The recovery of an unknown signal $\mu^\dagger$ comprising finitely many point sources lies at the heart of challenging applications such as acoustic or seismic inversion, optical imaging, signal processing and initial value identification. A popular mathematical model for the recovery of the locations $y_n^\dagger \in \Omega_s$ and amplitudes $q_n^\dagger$ of its $N_s^\dagger$ individual point sources is given by integral equations

$$z_j^d = \int_{\Omega_s} k(x_j, y) \,\mathrm{d}\mu^\dagger(y) + \varepsilon_j = \sum_{n=1}^{N_s^\dagger} q_n^\dagger k(x_j, y_n^\dagger) \quad \text{for } j = 1, \ldots, N_o. \tag{0.1}$$

Here, $k \in \mathcal{C}^2(\Omega_o \times \Omega_s)$ and $x_j \in \Omega_o$ denote a given integral kernel and measurement locations, respectively.

This type of *ill-posed inverse problem* is challenging for a variety of reasons. First and foremost, we neither assume knowledge on the amplitudes and positions of the sources nor on their number. This adds a combinatorial component to the, potentially highly nonlinear, problem. Second, inference on $\mu^\dagger$ is only possible through a finite number of indirect measurements z^d. Additional problems are posed by the appearance of unobservable, deterministic or random, measurement noise ε, in the problem. A recently popularized approach to alleviating many of these difficulties is to identify $\mu^\dagger$ with a finite linear combination of Dirac Delta functionals

$$\mu^\dagger = \sum_{n=1}^{N_s} q_n^\dagger \delta_{y_n^\gamma} \quad \text{where} \quad \int_\Omega k(x_j, y) \,\mathrm{d}\delta_{y_n^\dagger}(y) = k(x_j, y_n^\dagger). \tag{0.2}$$

Subsequently, we try to recover $\mu^\dagger$ by the stable solution of the linear, ill-posed, operator equation

$$\text{find } \mu \in \mathcal{M}(\Omega_s)\colon \quad z^d = K\mu \text{ where } K\mu = \left(\int_\Omega k(x_1, y) \,\mathrm{d}\mu(y), \ldots, \int_\Omega k(x_{N_o}, y) \,\mathrm{d}\mu(y) \right)$$

over the space of Radon measures $\mathcal{M}(\Omega_s)$ on Ω_s. At first glance, this might seem counter-intuitive: The space $\mathcal{M}(\Omega)$ is way larger than the set of "sparse" signals of the form (0.2). Thus, this lifting should contribute to the ill-posedness of the problem. However, it also bypasses the nonlinear dependency of $k(x_j, \cdot)$ onto the location of the sources and enables the use of powerful tools from variational regularization theory for the reconstruction of $\mu^\dagger$. Central objects in this context, are the (noiseless) *minimum Radon norm problem*

$$\min_{\mu\in\mathcal{M}(\Omega_s)} \|\mu\|_{\mathcal{M}(\Omega_s)} \quad s.t. \quad K\mu = K\mu^\dagger \tag{$\mathcal{P}_0$}$$

as well as the question whether $\mu^\dagger$ is *identifiable*, i.e. its unique solution. A sufficient condition for the latter, is the injectivity of the restricted operator $K_{|\operatorname{supp}\mu^\dagger}$ as well as the existence of a, in some sense minimal, dual certificate $\bar{\eta} \in \mathcal{C}^2(\Omega_s)$ satisfying a *strengthened source condition*

$$|\eta^\dagger(y)| \le 1 \quad \text{for all } y \in \Omega_s,\ \eta^\dagger(y_n^\dagger) = \operatorname{sign}(q_n^\dagger), \quad |\eta^\dagger(y)| < 1 \quad \text{for all } y \in \Omega_s \setminus \{y^\dagger\}_{n=1}^{N_s}.$$

For example, for a specific kernel k, the groundbreaking paper [1] shows that $\mu^\dagger$ is identifiable if the source locations $y_n^\dagger$ are sufficiently well separated.
However, measurements stemming from experiments are always affected by errors, either due to external influences, imperfectness of the measurement devices or human failure. These have to be taken into account in order to guarantee a stable recovery of $\mu^\dagger$. Despite the popularity of sparse inverse problems, all existing work, to the best of our knowledge, focuses on deterministic noise ε. In this context, several manuscripts, see e.g. [2], study the approximation of an identifiable $\mu^\dagger$ by solutions to the Tikhonov-regularized problem

$$\bar{\mu}(\varepsilon) \in \mathcal{M}(\varepsilon) \coloneqq \operatorname*{arg\,min}_{\mu\in\mathcal{M}(\Omega_s)} \left[\frac{1}{2}\|K\mu - z^d(\varepsilon)\|^2_{\Sigma_0^{-1}} + \beta\|\mu\|_{\mathcal{M}(\Omega_s)}\right], \tag{$P_{\beta,\varepsilon}$}$$

where Σ_0 is positive definite and the regularization parameter $\beta = \beta(\|\varepsilon\|)$ is adapted to the strength of the noise. This represents a challenging *nonsmooth* minimization problem over the infinite dimensional, non-reflexive, space of Borel measures. Moreover, due to its lack of strict convexity, its solutions are typically non-unique. Under mild conditions on the choice of β, arbitrary solutions $\bar{\mu}(\varepsilon)$ approximate $\mu^\dagger$ in the weak*-sense, i.e.

$$\int_{\Omega_s} \varphi(y)\ \mathrm{d}\bar{\mu}(\varepsilon)(y) \to \int_{\Omega_s} \varphi(y)\ \mathrm{d}\mu^\dagger(y) \quad \text{for all } \varphi \in \mathcal{C}(\Omega)$$

as ε goes to zero. Moreover, if the minimal dual certificate $\bar{\eta}$ associated to Problem ($\mathcal{P}_0$) satisfies the strengthened source condition and its curvature does not degenerate around $y_n^\dagger$, $\bar{\mu}(\varepsilon)$ is unique and of the form

$$\bar{\mu}(\varepsilon) = \sum_{n=1}^{N_s^\dagger} \bar{q}_n(\varepsilon)\delta_{\bar{y}_n(\varepsilon)} \quad \text{with} \quad |\bar{q}_n(\varepsilon) - q_n^\dagger| + |\bar{y}_n(\varepsilon) - y_n^\dagger| = \mathcal{O}(\|\varepsilon\|)$$

provided that $\|\varepsilon\|$ and β are small enough. Such *non-degeneracy conditions* for dual certificates have proven to be a key point in the analysis of sparse minimization problems, from their numerical solution, [6], to their optimal discrete approximation, [4].
While the literature on the deterministic case is rich, assuming knowledge on the norm of the error is very restrictive or even unrealistic from a practical point of view. In this talk, we therefore adopt a frequentistic viewpoint on sparse inverse problems and assume that the measurement noise follows a known Gaussian distribution $\varepsilon \sim \mathcal{N}(0, p^{-1}\Sigma_0)$ where Σ_0 is as above with $\operatorname{tr}(\Sigma_0^{-1}) = 1$ and $p > 0$ denotes the overall precision of the measurement error. Similar to the deterministic case, we rely on a Tikhonov-type estimator $\bar{\mu}(\varepsilon) \in \mathcal{M}(\varepsilon)$ and investigate its “closedness” to the ground truth $\mu^\dagger$. However, the randomness of the noise poses various new challenges but also offers previously unknown possibilities. First and foremost, we have to define a notion of “distance” between the estimator $\bar{\mu}$ and the ground truth $\mu^\dagger$. In this context, different to the

deterministic setting, our analysis cannot rely on smallness assumptions on the euclidean norm of the noise: Albeit with small probability, some realizations of ε might be very large. In this situation, solutions to ($P_{\beta,\varepsilon}$) can exhibit undesirable features such as clustering phenomena around $y_n^\dagger$ or spurious sources far away from the true support. In particular, the reconstructed signal may not consist of exactly $N_s^\dagger$ Dirac Delta functionals. Thus, we require a suitable distance $d(\cdot,\cdot)$ on $\mathcal{M}(\Omega_s)$ to compare two measures of, potentially, different support size. Note that the naive choice of $d(\cdot,\cdot) = \|\cdot - \cdot\|_{M(\Omega_s)}$ is not suitable since

$$\|q_1\delta_{y_1} - q_2\delta_{y_2}\|_{M(\Omega_s)} = |q_1| + |q_2| \quad \text{whenever} \quad y_1 \neq y_2.$$

Secondly, any measure of closedness between $\bar{\mu}$ and $\mu^\dagger$ should respect the random nature of the noise as well as its distribution. For this purpose, we advocate the use of the mean-squared error

$$\mathrm{MSE}[\bar{\mu}] = \mathbb{E}_\varepsilon \left[\sup_{\bar{\mu}(\varepsilon)\in\mathcal{M}(\varepsilon)} d_{\mathrm{HK}}(\bar{\mu}(\varepsilon), \mu^\dagger)^2 \right]$$

where $d_{\mathrm{HK}}(\cdot,\cdot)$ denotes a generalization of the Hellinger-Kantorovich distance, [5], to signed measures. In contrast to the Radon norm, $d_{\mathrm{HK}}(\cdot,\cdot)$ is compatible with the weak* topology on bounded subsets of $\mathcal{M}(\Omega_s)$.
Within this talk, we try to answer two questions. On the one hand, we discuss the asymptotic behavior of $\bar{\mu}$ in the high precision limit, i.e.:

1. If $p > 0$ is large enough and β is chosen appropriately, is $\mathrm{MSE}[\bar{\mu}]$ small?

Secondly, recall that we are given the whole distribution of the measurement data, not only one realization. Hence, we can try to mitigate the influence of the measurement errors on the estimator, *a priori*, i.e. before any measurements are taken, by optimizing the measurement setups:

2. For fixed $p > 0$ large enough, can we choose x_j and Σ_0 such that $\mathrm{MSE}[\bar{\mu}]$ is as small as possible?

Note that the motivation behind this second question, similar to von Mises work, is not a purely mathematical one but stems from the underlying application. Collecting experimental data is often associated to a substantial financial cost. Moreover, some experiments can only be carried out once since, e.g., some of the involved materials are destroyed irreversibly. Hence, planning the experiment a priori to ensure good results, and to save money, is greatly desired. Of course, formulating 2. directly as mathematical program leads to a computationally prohibitive problem, first, due to the difficult definition of d_{HK} and, second, due to the involved expectation. However, we still answer both questions, positively, showing that for some β_0 large enough and $\beta(p) = \beta_0/\sqrt{p}$, there holds

$$\mathbb{E}_\varepsilon \left[\sup_{\bar{\mu}(\varepsilon)\in\mathcal{M}(\varepsilon)} d_{\mathrm{HK}}(\bar{\mu}(\varepsilon), \mu^\dagger)^2 \right] \leq C_1 \frac{\Psi(\mathbf{x}, \Sigma_0)}{p} + C_2 \exp(-\beta_0^2/N_o).$$

for some $C_1,\ C_2 > 0$, which are essentially independent of $\mathbf{x}$ and Σ_0, and an easy-to-optimize design criterion $\Psi(\mathbf{x}, \Sigma_0)$ depending on the measurement setup. The latter dominates the overall error for β_0 and p large enough and thus provides a useful way to optimize the sensor setup.

The talk is accompanied by extensive numerical examples which, on the one hand, show the sharpness of the estimate and provide a justification for considering the easy-to-compute surrogate instead of the mean-squared-error based on the Hellinger-Kantorovich distance. On the other hand, we also provide some counterexamples which highlight the necessity of a proper choice for the measurement setup.

References

[1] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of super-resolution*, Commun. Pure Appl. Math., 67 (2014), pp. 906–956.

[2] V. Duval and G. Peyré, *Exact support recovery for sparse spikes deconvolution*, Found. Comput. Math., 15 (2015), pp. 1315–1355.

[3] P.-T. Huynh, K. Pieper, and D. Walter, *Towards optimal sensor placement for inverse problems in spaces of measures.* in preparation.

[4] D. Leykekhman, B. Vexler, and D. Walter, *Numerical analysis of sparse initial data identification for parabolic problems*, ESAIM, Math. Model. Numer. Anal., 54 (2020), pp. 1139–1180.

[5] M. Liero, A. Mielke, and G. Savaré, *Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures*, Invent. Math., 211 (2018), pp. 969–1117.

[6] K. Pieper and D. Walter, *Linear convergence of accelerated conditional gradient algorithms in spaces of measures*, ESAIM, Control Optim. Calc. Var., 27 (2021), p. 37. Id/No 38.