OPEN-ENDED CONFIGURATIONS OF RADIO TELESCOPES: TOWARDS OPTIMAL DESIGN

VLADIK KREINOVICH, SCOTT A. STARKS and OLGA KOSHELEVA
NASA Pan-American Center for Earth and Environmental Studies
University of Texas at El Paso
El Paso, TX 79968, USA
emails sstarks@utep.edu, vladik@cs.utep.edu, olga@ece.utep.edu

ANDREI FINKELSTEIN
Institute of Applied Astronomy
Russian Academy of Sciences
197042 St. Petersburg, Russia

ABSTRACT

The quality of radio astronomical images drastically depends on where we place the radio telescopes. During the design of the Very Large Array, it was empirically shown that the power law design, in which n-th antenna is placed at a distance n^β from the center, leads to the best image quality. In this paper, we provide a theoretical justification for this empirical fact.

KEYWORDS: Radio telescopes, Very Large Array, optimization under uncertainty

INTRODUCTION

Why radio telescopes

According to modern physics, most elementary particles are photons, i.e., quanta of electromagnetic field. Not surprisingly, the main information about the extra-terrestrial objects comes from observing electromagnetic waves on different wavelengths. The Earth's atmosphere absorbs most of these waves, so there are only a few windows of observability.

The most well known window corresponds to visible light. The corresponding optical telescopes indeed bring a lot of astronomical information. However, this information is often not sufficient: many celestial objects are not bright in visible light. To complement this information, astronomers use radio telescopes, devices that use the second observability window of radio waves.

Why configurations of radio telescopes

According to optics, when we use a telescope of diameter d to make observations on wavelength λ, we can determine the location of the radiation sources with an error $\approx \lambda/d$. To increase the observation quality, we must decrease this error, and thus, we must increase the diameter d. For radio telescopes, from the technical viewpoint, the largest possible diameter is ≈ 100 m. Thus, if we want to further increase the diameter d, we cannot simply design a *single* telescope of larger diameter. Instead, we must build a *configuration* of radio telescopes.

Why open-ended configurations of radio telescopes

In principle, the more telescopes we add, the more the noise decreases and therefore, the better the quality of the resulting images. However, telescopes are very costly devices, and these financial considerations severely limit our design abilities.

Sometimes, when a configuration is built, it turns out that for some observations, adding one or several appropriately placed radio telescopes would drastically increase the amount of astrophysical information that can be extracted from the resulting images. In this case, it makes sense to add a few telescopes to the existing configuration. In view of this possibility, many configurations are designed as *open-ended*, when it is always possible to add one or several telescopes.

We need optimal configurations

The image quality drastically depends on where exactly we place the telescopes. Depending on where we place them, we can get almost an order of magnitude improvement or decrease in image quality. We want to extract as much information from our investment in a radio telescope configuration as possible. Since telescopes are expensive, it makes sense to spend as much computational time and resources as necessary and find the truly optimal design.

Empirical analysis and the Very Large Array

The problem of optimally designing a configuration of radio telescopes was first handled during the design of the Very Large Array [1,2,4,5]. First, experimental and theoretical analysis showed that in the optimal open-ended design, radio telescopes are placed along several semi-lines with a common origin. If we select n lines, then each line should form an angle of $2\pi/n$ with the neighboring one. For example, if we select 3 lines, they form a Y-shape configuration; if we select $n = 4$, we get a cross-shaped configuration, etc.

For each such configuration, it is important to describe where exactly the antennas should be placed on each line. When we have a large number of telescopes, then we can describe the desired placement by describing the placement "density" $\rho(r)$, i.e., by describing, for each distance r from the origin, how many telescopes are placed at distance $\approx r$. So, to find the optimal placement, we must find the optimal placement density function $\rho(r)$.

Empirical comparison of several possible placement functions showed that for several different criteria, a power law $\rho(r) = C \cdot r^{-\alpha}$ leads to the best image quality [2,4,5]. This placement corresponds to placing n-th antenna at a distance $\sim n^{\beta}$ from the origin, where $\beta > 0$ is some constant related to α. Because of this analysis, this placement was selected for the design of VLA [2,4,5].

For some criteria, it was even possible to theoretically prove that this placement is optimal [1] – but, alas, not for the value α used in the actual VLA design. In this paper, we provide a theoretical proof that the power law placement is indeed optimal under any optimality criterion that satisfies certain reasonable properties.

TOWARDS MATHEMATICAL FORMULATION OF THE PROBLEM

Only Smooth Functions $\rho(r)$ Make Sense

The density function $\rho(r)$ is usually obtained by "smoothing" of the discrete distribution, so it makes sense only to consider smooth functions $\rho(r)$.

We Must Choose a Family of Functions, Not a Single Function

The number of telescopes depends on how we count. Since it is difficult to make a single perfect surface, some telescopes (called "active") consist of several independently moving pieces whose relative positions and/or connections are constantly being adjusted so that the total shape remains close to perfect. In this case, we can either count this telescope as one, or we can count it by the total number of its component antennas. If we change the way we count telescopes, then instead of the original density function $\rho(r)$, we get a new density function $C \cdot \rho(r)$, where C is the average number of components per originally counted telescope.

Thus, the functions $\rho(r)$ and $C \cdot \rho(r)$ describe exactly the same configuration. Hence, we cannot select a unique function $\rho(r)$ and claim it to be the best, because for every function $\rho(r)$, the function $C \cdot \rho(r)$ describes exactly the same configuration. In view of this, instead of formulating a problem of choosing the best density *function*, it is more natural to formulate a problem of choosing the best *family* $\{C \cdot \rho(r)\}_C$ of density functions.

Which Family Is the Best? We May Need Non-Numerical Optimality Criteria

Among all the families $\{C \cdot \rho(r)\}_C$, we want to choose the best one.

In mathematical optimization problems, numerical criteria are most frequently used, when to every alternative (in our case, to each family) we assign some value expressing its performance, and we choose an alternative (in our case, a family) for which this value is the largest. In our problem, as such a numerical criterion, we can select, e.g., the average approximation error A, measured as the mean square deviation between the image reconstructed by the corresponding configuration and the original image.

However, it is not necessary to restrict ourselves to such numerical criteria only. For example, if we have several different families that have the same average approximation error A, we can choose between them the one that has the minimal computation time T for image reconstruction. In this case, the actual criterion that we use to compare two families is not numerical, but more complicated: *a family F_1 is better than the family F_2 if and only if either $A(F_1) < A(F_2)$, or $A(F_1) = A(F_2)$ and $T(F_1) < T(F_2)$*. A criterion can be even more complicated. What a criterion *must* do is to allow us, for every pair of families, to tell whether the first family is better with respect to this criterion (we'll denote it by $F_1 \succ F_2$), or the second is better ($F_1 \prec F_2$), or these families have the same quality in the sense of this criterion (we'll denote it by $F_1 \sim F_2$). Of course, it is necessary to demand that these choices be consistent, e.g., if $F_1 \prec F_2$ and $F_2 \prec F_3$ then $F_1 \prec F_3$.

The Optimality Criterion Must Select a Unique Optimal Family

Another natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family). The reason for this demand is very simple.

If a criterion does not choose a family at all, then it is of no use.

If several different families are "the best" according to this criterion, then we still have a problem to choose among those "best". Therefore, we need some additional criterion for that choice. For example, if several families turn out to have the same average approximation error, we can choose among them a with the minimal computation time.

So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion or according to the old criterion they had the same quality and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family it means that this criterion is not final. We have to modify it until we come to a final criterion that will have that property.

The Optimality Criterion Must Be Scale-Invariant

The next natural condition that the criterion must satisfy is connected with the fact that the numerical value of the distance r depends on the choice of the unit for measuring distance.

If we replace the original unit of length by a new unit which is λ times larger (i.e., replace feet by meters), then numerical values change from r to $\widetilde{r} = r/\lambda$. How will the density function $\rho(r) = N([r_1, r_2])/(r_2 - r_1)$ ($r_i \approx r$) look in the new units?

Let us assume that in the new units, the distance between the two points equals $\widetilde{r}$. Then, the same distance in the old units is equal to $r = \lambda \cdot \widetilde{r}$. Thus, the new density function is equal to

$$\widetilde{\rho}(\widetilde{r}) = \frac{\widetilde{N}([\widetilde{r}_1, \widetilde{r}_2])}{\widetilde{r}_2 - \widetilde{r}_1} = \lambda \cdot \frac{N([\lambda \cdot \widetilde{r}_1, \lambda \cdot \widetilde{r}_2])}{\lambda \cdot \widetilde{r}_2 - \lambda \cdot \widetilde{r}_1} = \lambda \cdot \rho(\lambda \cdot \widetilde{r}).$$

So, the configuration that in the old units is described by a family $\{C \cdot \rho(r)\}$, in the new units, has a new form $\{C \cdot \lambda \cdot \rho(\lambda \cdot r)\}$, i.e., equivalently, $\{C \cdot \rho(\lambda \cdot r)\}$,

Since this change is simply a change in a unit of length, it is reasonable to require that going from $\rho(r)$ from $\rho(\lambda \cdot r)$ should not change the *relative* quality of the density functions, i.e., if a family $\{C \cdot \rho(r)\}_C$ is better that the family $\{C \cdot \rho'(r)\}_C$, then for every $\lambda > 0$, the family $\{C \cdot \rho(\lambda \cdot r)\}_C$ must be still better than the family $\{C \cdot \rho'(\lambda \cdot r)\}_C$.

So, we arrive at the following definitions.

DEFINITIONS AND THE MAIN RESULT

Definition 1.

- *By a density function we mean a smooth function from non-negative real numbers to non-negative real numbers.*
- *By a family of functions we mean the family $\{C \cdot \rho(r)\}_C$, where $\rho(r)$ is a given density function and C runs over arbitrary positive real numbers.*
- *A pair of relations $(\prec, \sim)$ is called consistent* [3] *if it satisfies the following conditions:*

  (1) *if $a \prec b$ and $b \prec c$ then $a \prec c$;*

  (2) *$a \sim a$;*

  (3) *if $a \sim b$ then $b \sim a$;*

  (4) *if $a \sim b$ and $b \sim c$ then $a \sim c$;*

  (5) *if $a \prec b$ and $b \sim c$ then $a \prec c$;*

  (6) *if $a \sim b$ and $b \prec c$ then $a \prec c$;*

  (7) *if $a \prec b$, then $b \prec a$ or $a \sim b$ are impossible.*

Definition 2.

- *Assume a set A is given. Its elements will be called alternatives. By an optimality criterion we mean a consistent pair $(\prec, \sim)$ of relations on the set A of all alternatives. If $b \prec a$, we say that a is better than b; if $a \sim b$, we say that the alternatives a and b are equivalent with respect to this criterion.*
- *We say that an alternative a is optimal (or best) with respect to a criterion $(\prec, \sim)$ if for every other alternative b either $b \prec a$ or $a \sim b$.*
- *We say that a criterion is final if there exists an optimal alternative, and this optimal alternative is unique.*
- *Let $\lambda > 0$ be a real number. By the λ-rescaling $R_\lambda(\rho)$ of a function $\rho(r)$ we mean a function $(R_\lambda(\rho))(r) \stackrel{\text{def}}{=} \rho(\lambda \cdot r)$.*
- *By the λ-rescaling $R_\lambda(F)$ of a family F, we mean the set of the functions that are obtained from $f \in F$ by λ-rescaling.*

In this paper, we consider optimality criteria on the set $\mathcal{F}$ of all families.

Definition 3. *We say that an optimality criterion on F is scale-invariant if for every two families F and G and for every number $\lambda > 0$, the following two conditions are true:*

i) *if F is better than G in the sense of this criterion (i.e., $G \prec F$), then*

$$R_\lambda(G) \prec R_\lambda(F);$$

ii) *if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then*

$$R_\lambda(F) \sim R_\lambda(G).$$

As we have already remarked, the demands that the optimality criterion is final and scale-invariant are quite reasonable. The only problem with them is that at first glance they may seem rather weak. However, they are not, as the following Theorem shows:

Theorem. [3] *If a family F is optimal in the sense of some optimality criterion that is final and scale-invariant, then every density function $\rho(r)$ from this optimal family F which has the form $\rho(r) = A \cdot r^{-\alpha}$ for some real numbers A and α.*

So, the configuration used in the VLA design in indeed optimal.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209 and grant NCC 2-1232, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grants numbers F49620-95-1-0518 and F49620-00-1-0365, by grant No. W-00016 from the U.S.-Czech Science and Technology Joint Fund, by NSF grants CDA-9522207, ERA-0112968 and 9710940 Mexico/Conacyt, and by IEEE/ACM SC2001 Minority Serving Institutions Participation Grant.

REFERENCES

1. Chow, Y. L. “On designing a supersynthesis antenna array”, *IEEE Transactions on Antennas and Propagation* AP-20 (1972) 30–35.

2. Napier, P. J., Thompson, A. R., and Ekers, R. D. “The Very Large Array: design and performance of a modern synthesis radio telescope”, *Proc. IEEE* 71 (1983) 1295–1320.

3. Nguyen, H. T., and Kreinovich, V., *Applications of continuous mathematics in computer science* (Kluwer, Dordrecht, 1997).

4. Thompson, A. R., Clark, B. G., Wade, C. M., and Napier, P.J. “The Very Large Array”, *Astrophys. J. Suppl.* 44 (1980) 151–167.

5. Thompson, A. R., Moran, J. M., Swenson Jr., G. W. *Interferometry and Synthesis in Radio Astronomy* (Wiley, New York, 2001).