OULU 2012
C 424

ACTA UNIVERSITATIS OULUENSIS

Davide Macagnano

MULTITARGET LOCALIZATION AND TRACKING

ACTIVE AND PASSIVE SOLUTIONS

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU, FACULTY OF TECHNOLOGY,
DEPARTMENT OF COMMUNICATIONS ENGINEERING;
CENTRE FOR WIRELESS COMMUNICATIONS;
INFOTECH OULU

ACTA UNIVERSITATIS OULUENSIS
C Technica 424

DAVIDE MACAGNANO

MULTITARGET LOCALIZATION AND TRACKING

Active and passive solutions

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in OP-sali (Auditorium L10), Linnanmaa, on 27 June 2012, at 9 a.m.

UNIVERSITY OF OULU, OULU 2012

Acta Univ. Oul. C 424, 2012

Supervised by
Professor Giuseppe Abreu
Professor Jari Iinatti

Reviewed by
Professor Robert Piché
Associate Professor Henk Wymeersch

ISBN 978-951-42-9858-5 (Paperback)
ISBN 978-951-42-9859-2 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2012

Macagnano, Davide, Multitarget localization and tracking. Active and passive solutions

University of Oulu Graduate School; University of Oulu, Faculty of Technology, Department of Communications Engineering; Centre for Wireless Communications; Infotech Oulu, P.O. Box 4500, FI-90014 University of Oulu, Finland

Acta Univ. Oul. C 424, 2012
Oulu, Finland

Abstract

Localization and tracking of multiple targets is becoming an essential feature of modern communication services and systems. Although necessary in many contexts, such as surveillance and monitoring applications, low-complexity and reliable solutions capable of coping with different degrees of information are not yet available.

This thesis deals with different problems that are encountered in localization and tracking applications and aims to establish a broad understanding of multitarget systems ranging from complete active to incomplete passive solutions in dynamic scenarios. Thereby we start by investigating a fully algebraic framework which is proved to be advantageous in dynamic contexts characterized by no *a-priori* knowledge. Subsequently we extend the approach to improve its robustness versus corrupted observations. Finally we focus on a Bayesian formulation of the passive multitarget tracking (MTT) problem.

The Thesis is based on three parts. The first part focuses on a low complexity mathematical representation of the active problem (*i.e* manifold-based solution). In particular, the spectrum of the matrices used to represent target locations within an algebraic, multidimensional scaling (MDS) based, solution is characterized statistically. In so doing we propose a novel Jacobi-based eigenspace tracking algorithms for Gramian matrices which is shown to be particularly convenient in a multidimensional scaling formulation of the multitarget tracking problem.

The second part deals with incomplete-active multitarget scenarios as well as eventual disturbances on the ranging measurements such as bias due to non-line-of-sight conditions. In particular the aforementioned algebraic solution is extended to cope with heterogeneous information and to incorporate eventual knowledge on the confidence of the measurement information. To do so we solve the classical multidimensional scaling (C-MDS) over a novel kernel matrix and show how the intrinsic nature of this formulation allows to deal with heterogeneous information, specifically angle and distance measurements.

Finally, the third part focuses on the *random finite sets* formulation of Bayesian multisensor MTT problem for passive scenarios. In this area a new gating strategy is proposed to lower the computational complexity of the algorithms without compromising their performance.

Keywords: adaptive gating, Jacobi algorithm, multidimensional scaling (MDS), multitarget localization, passive multitarget tracking, subspace tracking

Macagnano, Davide, Monikohdepaikannus ja seuranta. Aktiiviset ja passiiviset ratkaisut
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Teknillinen tiedekunta, Tietoliikennetekniikan osasto; Centre for Wireless Communications; Infotech Oulu, PL 4500, 90014 Oulun yliopisto
Acta Univ. Oul. C 424, 2012
Oulu

Tiivistelmä

Useiden kohteiden yhtäaikaisesta paikannuksesta ja seurannasta on tulossa olennainen osa nykyaikaisia viestinnän palveluita ja järjestelmiä.

Huolimatta siitä, että yhtäaikainen paikannus on erittäin tarpeellinen osa monissa yhteyksissä, kuten valvonnan ja kontrolloinnin sovelluksissa, siihen ei ole olemassa kompleksisuudeltaan alhaista ratkaisua, joka ottaisi huomioon kaiken saatavilla olevan informaation.

Väitöskirja käsittelee useiden kohteiden paikannukseen ja seurantaan liittyviä ongelmia, ja se keskittyy antamaan laajan ymmärryksen aktiivisista täydellisistä menetelmistä passiivisiin epätäydellisiin menetelmiin dynaamisissa ympäristöissä. Saavuttaakseen tavoitteen väitöskirjassa esitetään algebrallinen kehys, jonka todistetaan olevan edistyksellinen dynaamisissa ympäristöissä, joissa ei ole ennakkoinformaatiota saatavilla. Seuraavaksi väitöskirja laajentaa esitettyä lähestymistapaa parantamalla sen vakautta vääriä havaintoja vastaan. Lopuksi esitetään bayesialainen formulointi passiiviselle usean kohteen seuranta -ongelmalle (MTT).

Väitöskirja on jaettu kolmeen on osaan. Ensimmäinen osa käsittelee aktiivisen ongelman kuvaamista matemaattisesti säilyttäen alhaisen kompleksisuuden. Erityisesti tässä osassa karakterisoidaan tilastollisesti matriisien spektrin käyttäminen kohteiden paikan esittämiseen moniulotteiseen skaalaukseen (MDS) pohjautuvassa menetelmässä. Saavuttaakseen tämän väitöskirja esittää Jacobin ominaisavaruuksiin perustuvan seuranta-algoritmin Gramian matriiseille, joiden osoitetaan olevan erityisen soveltuvia usean kohteen seuraamisongelman kuvaamiseen MDS-menetelmän avulla.

Toinen osa käsittelee epätäydellistä aktiivista usean kohteen skenaariota, kuten myös mittausten lopullisia häiriötä, esim. ei-näköyhteyskanavasta johtuvaa harhaa. Edellä mainittu algebrallinen ratkaisu on laajennettu ottamaan huomioon heterogeeninen informaatio sekä tieto mittausdatan luotettavuudesta. Lisäksi tässä osassa esitetään ratkaisu klassiseen moniulotteiseen skaalausongelmaan (C-MDS) esittelemällä uudenlainen ydinmatriisi ja osoitetaan, kuinka tämä mahdollistaa heterogeenisen informaation, tässä tapauksessa kulma-ja etäisyysmittauksien, huomioon ottamisen.

Viimeisessä osassa käsitellään äärellisten satunnaisten joukkojen soveltuvuutta bayesialaisen MTT-ongelman ratkaisuun passiivisissa skenaarioissa. Väitöskirja esittää uuden porttistrategian algoritmien kompleksisuuksien pienentämiseksi säilyttäen kuitenkin samalla niiden suorituskyvyn.

Asiasanat: adaptiivinen portitus, alivaruusseuranta, Jacobin algoritmi, monikohdepaikannus, moniulotteinen skaalaus (MDS), passiivinen monikohdeseuranta

To my brother Daniele

Acknowledgements

The research work presented in this thesis was carried out at the Centre for Wireless Communications (CWC), University of Oulu, Finland. This has been possible thanks to the support of several persons, some of them I wish to acknowledge personally.

Giuseppe (Abreu), I still remember when just arrived at Oulu's airport you already started engaging me in your plans for our future research. Since then I had the opportunity to experience the incredible enthusiasm and dedication you put into your research work, and I am grateful you always had time to discuss new ideas with me too. I grew considerably during these years and if now I can consider myself a self-critical and committed researcher it is thanks to your efforts and your example. In addition to the professional side, I am grateful for having shared these years with a very good friend with whom I spent some memorable times, like our skiing experiences in Auranmaja or the fishing attempts on Kiiminki river.

Professor Jari Iinatti, I am very grateful for the help you gave me during the redaction of my Thesis. Your support and comments significantly improved the quality of this manuscript. I want to thank the reviewers A.-Professor Henk Wymeersch from Chalmers University, Sweden and Professor Robert Pichè from Tampere University of Technology, Finland, whose constructive criticisms helped me to improve the clarity of this work.

My gratitude goes also to Professor Matti Latva-aho and Lic. Tech. Ari Pouttu, the directors of the CWC during my years here, for providing an energetic and inspiring working environment.

Giuseppe (Destino). On the way towards a Ph.D. it is normal to encounter difficult times. However, to be able to count on an excellent colleague and a good friend makes everything much easier. I am grateful for the countless technical discussions we had and what I learnt from them. Thank you Jani for the good time we share at the office and for helping me translate the Finnish part of this thesis. I would also like to thank all my office mates, with whom I had the pleasure to share my lunch breaks and ski trips and many other activities, Alberto, Leonardo, Antti, Markus, Jari, Esa, Carlos, Stefano, Jussi, Pedro,

Brett as well as Professor Behnaam Aazhang. My gratitude goes as well to the administrative staff that worked at the CWC during these years, Hanna, Kirsi, Timo, Elina and Sari.

The thesis work has been financially supported by GETA Graduate School during the period of 2008-2012, and through the European projects PULSERS II, EUWB and BUTLER, which I highly acknowledge. I was also fortunate to receive personal grants for doctoral studies from the following Finnish foundations: Nokia Oyj:n säätiö, Oulun yliopiston tukisäätiö, ELISA-säätiö, Riitta ja Jorma J. Takasen säätiö, and Walter Ehrström-säätiö. These contributions stimulated me to engage even more in my research work.

A special thank goes to my wife Virve for having been so patient during all this time. Without your support and understanding everything would have been much harder. Thank you my little Elena for allowing me ten extra days from the original deadline to complete the manuscript, not sure mum appreciated as much as I did though.

Se questa esperienza ha avuto un inizio lo devo a te Andrea (Rinaldo) che quando si e' presentata l'opportunità di venire ad Oulu mi hai aiutato a fare la scelta giusta. Marco, Massimiliano, Giovanni, sono grato che la nostra amicizia sia più forte delle distanza che ci separa.

Un grazie speciale va a mio fratello Andrea e ai miei genitori per l'amore ed il sostegno che mi hanno sempre saputo dare e che continuano a darmi quotidianamente.

Symbols and abbreviations

$\angle(\cdot,\cdot)$	angle between two points
$\lvert\cdot\rvert$	absolute of the argument
$\lVert\cdot\rVert_{\mathrm{F}}$	Frobenius norm of the matrix
$(\cdot)^{\mathrm{T}}$	transpose of the argument
$(\cdot)^{\odot m}$	m-th element-wise (Hadamard) power
$(\cdot)^{\dagger}$	Moore-Penrose pseudo-inverse
$\breve{(\cdot)}$	approximation of the argument
$\tilde{(\cdot)}$	perturbation of the argument
$\hat{(\cdot)}$	estimate of the argument
$(\cdot)_{\bar{k}}$	argument at time $\bar{k}$
$[\cdot]_{ij}$	the (i,j)'th entry of the matrix given as argument
$[\cdot]_{a:b,c:d}$	elements from the a-th to the b-th rows and from the c-th to the d-th columns of the matrix given as argument
$\mathrm{rank}(\cdot)$	rank of the matrix given as argument
$\mathrm{diag}(\cdot)$	diagonal matrix with the elements of the matrix given as argument on the main diagonal
$\mathrm{trace}(\cdot)$	trace of the matrix given as argument
$\det(\cdot)$	determinant of the matrix given as argument
$\mathbf{Tria}(\cdot)$	lower triangular component of a general triangularization algorithm
$\sim$	distributed according to
$\approx$	approximately equal to
$\triangleq$	defined as
$\circ$	Hadamard product
$\oslash$	inverse of the Hadamard product
$\otimes$	Kronecker product
$\partial(\cdot)/\partial(\cdot)$	differential operator
$\underset{p}{\geqslant}$	indicates that a random variable is larger than another with probability p
$\underset{p}{\leqslant}$	indicates that a random variable is smaller than another with probability p

$\Pr\{\cdot\}$	probability of the random variable given as argument
$\mathrm{E}[\cdot]$	expectation
$\perp$	boolean indicator
$\log_{10}(\cdot)$	logarithm in base 10
$\cos(\cdot)$	cosine function
$\sin(\cdot)$	sine function
$\cosh(\cdot)$	hyperbolic cosine function
$\mathrm{atan}(\cdot)$	arctangent function
$\mathbf{0}$	zero matrix; a subscript can be used to indicate the dimension
$\mathbf{1}$	vector of ones; a subscript can be used to indicate the dimension
$\arg\max(\cdot)$	maximum argument
$\arg\min(\cdot)$	minimum argument
$\mathcal{N}(\mathbf{x};\mathbf{m},\mathbf{\Sigma})$	multidimensional Gaussian with mean $\mathbf{m}$ and covariance matrix $\mathbf{P}=\mathbf{\Sigma}\cdot(\mathbf{\Sigma})^{\mathrm{T}}$
$\mathbf{a}$	$\mathbb{R}^N$ centering vector
$\mathcal{A}^{(i)}$	area defined by the i-th pair of measurements
$\mathcal{A}_{\mathrm{T}}$	area defined by the intersection of $\mathcal{A}^{(i)},\ \forall i=\{1,\ldots,B_A\}$
$\boldsymbol{A}$	$\mathbb{R}^{B_A\times B_A}$ first minor of $\mathbf{K}_{\mathrm{E}}$
$\mathbf{A}$	$\mathbb{R}^{N\times N}$ component of $\mathbf{G}_T^*$
b_d	parameter used to simulate NLOS range observations
B_A	number of combinations between the anchor nodes
B_T	$M-B_A$
$\mathbf{B}$	$\mathbb{R}^{N\times N}$ component of $\mathbf{G}_T^*$
$\boldsymbol{B}_{\bar{k}}$	RFS describing the spewing targets at $\bar{k}$
c	cutoff value in the OSPA metric
$c_{n_{\boldsymbol{z}}}$	surface of the unit hypersphere in the $n_{\boldsymbol{z}}$ dimensional space
c_{p}	propagation velocity
C_j^{ℓ}	binomial coefficient
$\mathcal{C}(\cdot)$	cardinality operator
$\mathcal{C}_S(\cdot)$	STRESS cost function
$\mathcal{C}_T(\cdot)$	STRAIN cost function
$\mathcal{C}_M(\cdot)$	majorizing function of STRESS
$\boldsymbol{C}$	$\mathbb{R}^{M\times N}$ involutive operator
$\mathbf{C}$	$\mathbb{R}^{N\times\eta}$ component of $\mathbf{B}$
$\mathbf{C}_i$	$i\in\{1,\ldots,4\}$ constant terms used in Nyström

d_0	reference distance in RSS measurements
d_g	Mahalanobis distance
d_{ij}	Euclidean distance between the $\mathbf{x}_i$ and $\mathbf{x}_j$
$\bar{d}^{(c)}(\cdot,\cdot)$	Weierstrass distance
$\bar{d}_p^{(c)}(\cdot,\cdot)$	OSPA metric
$\mathbf{d}_{\mathrm{E}}$	$\mathbb{R}^M$ Euclidean length for the M edges in $\vec{\mathbf{V}}_{\mathrm{E}}$
$\mathbf{d}_M$	$\mathbb{R}^M$ Euclidean distance from the origin for the M edges in $G_{\eta,N}(\cdot,\cdot,\cdot)$
$\mathbf{d}_N$	$\mathbb{R}^N$Euclidean distance from the origin for the N points in $\mathbf{X}$
$\mathbf{d}_{B_A}$	$\mathbb{R}^{B_A}$ Euclidean length for the B_A edges in $\vec{\mathbf{V}}_{\mathcal{A}}$
$\mathbf{d}_{B_T}$	$\mathbb{R}^{B_T}$ Euclidean length for the B_T edges in $\vec{\mathbf{V}}_{\mathbf{T}}$
$D_\nu(x)$	Parabolic Cylinder function
$\mathcal{D}(\mathbf{X})$	functional relation between $\mathbf{X}$ and the corresponding EDM
$\mathbb{D}_i$	i-th Gershgorin disc of $\mathbf{K}$
$\mathbb{D}_N(\eta)$	space of EDM generated by N points in a η dimensional space
$\mathbf{D}$	$\mathbb{R}^{N\times N}$ EDM
$\tilde{\mathbf{D}}$	$\mathbb{R}^{N\times N}$ perturbed EDM
$e_j(\cdot)$	elementary symmetric function
$\exp(\cdot)$	exponential function
$\mathbb{E}_N$	closed convex set of possible dissimilarity matrices
$\boldsymbol{E}_{\bar{k}}$	RFS describing the prediction of the surviving targets at $\bar{k}$
$f(\cdot)$	dynamic model
$f_C(\cdot)$	generic cost function
$f_M(\cdot)$	majorizing function of $f_C(\cdot)$
$\mathcal{F}_{\tilde{\mathbb{G}}_q}$	indicator function for the superdisc $\tilde{\mathbb{G}}_q$
$F(\cdot)$	space of all finite subsets of space given as argument$\mathcal{X}$
${}_2F_0(\cdot,\cdot;\cdot;\cdot)$	Hypergeometric function
$\mathbf{F}_{\bar{k}-1}$	$\mathbb{R}^{n_{\boldsymbol{x}}\times n_{\boldsymbol{x}}}$ transition matrix
$g(\cdot)$	measurement model
$\mathcal{G}_Q$	set of Gershgorin groups
$G_{\eta,N}(\cdot,\cdot,\cdot)$	oriented graph
$\mathbb{G}_N$	subset of $\mathbb{S}_N$
$\mathbb{G}_k$	k-th Gershgorin group
$\boldsymbol{G}$	$\mathbb{R}^{n_{\boldsymbol{z}}\times n_{\boldsymbol{x}}}$ observation matrix
$\mathbf{G}$	$\mathbb{R}^{N\times N}$ Gram matrix

$\mathbf{G}_T$	$\mathbb{R}^{N\times N}$ Gram matrix at T
$\mathbf{G}_T^*$	$\mathbb{R}^{N\times N}$ rotated Gram matrix at T
$h_{\bar{k}}(\cdot\|\cdot)$	multitarget likelihood function at $\bar{k}$
$\boldsymbol{H}_{\bar{k}}$	RFS accounting for the clutter measurements at $\bar{k}$
$\mathbf{H}$	$\mathbb{R}^{N\times N}$ eigenvalue matrix for $\breve{\mathbf{K}}$
$\mathbf{H}_{\mathbf{W}}$	$\mathbb{R}^{N\times N}$ matrix accounting for $\mathbf{W}$ in $\mathcal{C}_M$
$\bar{i}$	iteration index
$\mathcal{I}$	set of indexes associated with the Gershgorin of $\mathbf{K}$
$\mathcal{I}(\cdot)$	Gaussian integral of the function given as argument
$I_0(\cdot)$	modified Bessel function of order 0
$\mathbf{I}$	identity matrix; a subscript can be used to indicate the dimension
$J_{\bar{k}}$	number of GM components at $\bar{k}$
$J_{\Gamma,\bar{k}}$	number of GM components at $\bar{k}$ for the RFS Γ
$\mathbf{J}$	$\mathbb{R}^{N\times N}$ double-center matrix
$\mathbf{J}^*$	$\mathbb{R}^{N\times N}$ centering matrix for dynamic scenarios
$\mathbf{J}(\mathbf{a})$	$\mathbb{R}^{N\times N}$ generic centering matrix
$\bar{k}$	discrete time index
$\mathcal{K}(\boldsymbol{\Delta},\mathbf{a})$	functional relation between the kernel matrix $\mathbf{K}_{\mathbf{a}}$, $\boldsymbol{\Delta}$ and $\mathbf{a}$
$\boldsymbol{K}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}}\times n_{\boldsymbol{z}}}$ Kalman gain
$\mathbf{K}$	$\mathbb{R}^{N\times N}$ double-centered kernel
$\mathbf{K}^*$	$\mathbb{R}^{N\times N}$ dynamic double-centered kernel
$\mathbf{K}_{\mathbf{a}}$	$\mathbb{R}^{N\times N}$ kernel matrix
$\mathbf{K}_{\mathrm{E}}$	$\mathbb{R}^{M\times M}$ Edge-kernel matrix
$\mathbf{K}_{(i)}$	$\mathbb{R}^{N\times N}$ kernel matrix centered at the i-th anchor matrix
L_{d_0}	attenuation at d_0
$L_{\bar{k}}$	number of targets at $\bar{k}$ in a MTT scenario
$L_{1/2}^{\nu}(\cdot)$	generalized Laguerre polynomial of degree 1/2
m_n	measurement noise
$\mathbf{m}_{\bar{k}}^{(j)}$	$\mathbb{R}^{n_{\boldsymbol{z}}}$ mean vector for the j-th component of a GM
$\mathbf{m}_{\Gamma,\bar{k}}^{(j)}$	$\mathbb{R}^{n_{\boldsymbol{x}}}$ mean vector for the j-th component of a GM representing the RFS **Gamma**
$\mathbf{m}_{\mathrm{S},\bar{k}}^{(j)}$	$\mathbb{R}^{n_{\boldsymbol{z}}}$ mean vector for the j-th component of a GM representing the surviving targets at $\bar{k}$
$\mathcal{M}_{q_k}(\cdot,\cdot)$	envelope moment function of the Nakagami-m pdf
M	number of combination without repetition between N points

$M_{\bar{k}}$	number of observations at $\bar{k}$ in a MTT scenario
$\boldsymbol{M}_{\bar{k}}$	RFS accounting for the detection probability p_{D}
$n_{\bar{k}}$	cardinality of the RFS $\mathbf{X}_{\bar{k}}$ at $\bar{k}$
$n_{\boldsymbol{x}}$	size of the state vector $\boldsymbol{x}$
$n_{\boldsymbol{z}}$	size of the observation vector $\boldsymbol{z}$
N	number of targets in the scenario
N_c	number of cubature points
N_{S}	number of sweeps for the Jacobi algorithm
$\mathcal{O}(\cdot)$	complexity order
$p(\cdot)$	probability density function of the argument
$p_{\bar{k}}(\cdot)$	probability density function of the argument at $\bar{k}$
$p_{\mathrm{C},\bar{k}\mid\bar{k}}(\cdot)$	cardinality distribution at $\bar{k}$
$p_{\Gamma,\bar{k}}(\cdot)$	cardinality distribution for the RFS Γ at $\bar{k}$
p_{D}	detection probability
$p_{K,\bar{k}}(\cdot)$	distribution intensity of clutter arrival at $\bar{k}$
p_{S}	survival probability
$p_{\mathrm{S},\bar{k}}(\cdot)$	cardinality distribution of the surviving targets at $\bar{k}$
$p_{\mathbf{S},ij}$	(i,j)'th entry of $\mathbf{P}_{\mathbf{S}}$
$p_{\mathrm{Gauss}}(\cdot;\cdot,\cdot)$	monodimensional Gaussian distribution
$p_{\mathrm{Gamma}}(\cdot;\cdot,\cdot)$	monodimensional Gamma distribution
$p_{\mathrm{Gumbel}}(\cdot;\cdot,\cdot)$	monodimensional Gumbel distribution
$p_{\mathrm{Nakagami}}(\cdot;\cdot,\cdot)$	monodimensional Nakagami distribution
$p_{\mathrm{NSC}\chi}(\cdot;\cdot,\cdot,\cdot)$	monodimensional non-central χ distribution
$p_{\mathrm{Rayleigh}}(\cdot;\cdot)$	monodimensional Rayleigh distribution
$p_{\mathrm{SC}\chi}(\cdot;\cdot,\cdot)$	monodimensional central χ distribution
$p_{\mathrm{Tikhonov}}(\cdot;\cdot)$	monodimensional Tikhonov distribution
P_g	gate's probability
P_{j+u}^{n}	permutation coefficient
P_{t}	transmitted power
$P_{\mathrm{Gamma}}(\cdot;\cdot,\cdot)$	Gamma cumulative density function
$\mathcal{P}(\mathcal{I})$	powerset of $\mathcal{I}$
$\mathbf{P}_{\mathbf{S}}$	$\mathbb{R}^{N\times N}$ matrix used in the SMACOF algorithm
$\mathbf{P}_{\boldsymbol{x},\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}}\times n_{\boldsymbol{x}}}$ state covariance matrix at $\bar{k}$
$\mathbf{P}_{\boldsymbol{xz},\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}}\times n_{\boldsymbol{z}}}$ cross-covariance matrix at $\bar{k}$
$\boldsymbol{q}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}}}$ process noise at $\bar{k}$

Q	number of Geshgorin groups
$\mathbf{Q}$	$\mathbb{R}^{N\times N}$ eigenvector matrix for $\breve{\mathbf{K}}$
$\boldsymbol{Q}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}}\times n_{\boldsymbol{x}}}$ covariance matrix for the process noise at $\bar{k}$
$\boldsymbol{r}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{z}}}$ measurement noise at $\bar{k}$
R	second parameter Rayleigh distribution
$\mathcal{R}(i,j,\vartheta)$	Jacobi-Givens rotation
$\boldsymbol{R}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{z}}\times n_{\boldsymbol{z}}}$ covariance matrix for the measurement noise at $\bar{k}$
$\mathbb{S}_0$	Gershgorin null superdisc
$\mathbb{S}_N(\eta)$	space of symmetric PSD matrices with rank at most η
$\mathbf{S}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{z}}\times n_{\boldsymbol{z}}}$ innovation covariance matrix at $\bar{k}$
$\mathcal{T}(\cdot)$	functional relation between the Gram matrix $\mathbf{G}$ and the EDM $\mathbf{D}$.
$\bar{T}$	reference time
T_g	gate's threshold
$\mathbf{T}$	$\mathbb{R}^{B_T\times B_T}$ second minor of $\mathbf{K}_{\mathrm{E}}$
$\mathcal{U}(\cdot,\cdot)$	monodimensional uniform distribution
$U(\cdot,\cdot;\cdot)$	Hypergeometric function also known as Kummer's function
$\mathbf{U}$	$\mathbb{R}^{\eta\times\eta}$ right singular vectors of $\mathbf{X}$
$\mathbf{U}_{\boldsymbol{A}}$	$\mathbb{R}^{B_A\times B_A}$ set of eigenvectors of $\boldsymbol{A}$
$\mathcal{V}(\cdot)$	functional relation between $\vec{\mathbf{V}}$ and $\mathbf{X}$
V	surveillance volume used during filtering
V_{S}	total surveillance volume
V_{U}	upper bound for the surveillance volume
$V_{\mathrm{A}}^{(j)}(\cdot)$	surveillance volume j-th adaptive gate
$V_{\mathrm{E}}^{(j)}(\cdot)$	surveillance volume j-th elliptical gate
$\mathbf{V}$	$\mathbb{R}^{N\times N}$ matrix containing the eigenvectors of $\mathbf{K}$
$\mathbf{V}_{\mathbf{a}}$	$\mathbb{R}^{N\times N}$ matrix containing the eigenvectors of $\mathbf{K}_{\mathbf{a}}$
$\vec{\mathbf{V}}$	$\mathbb{R}^{M\times M}$ set of edges
$\vec{\mathbf{V}}_{\boldsymbol{A}}$	$\mathbb{R}^{B_A\times B_A}$ set of edges for $\boldsymbol{A}$
$\vec{\mathbf{V}}_{\mathrm{E}}$	$\mathbb{R}^{M\times M}$ set of edges for $\mathbf{K}_{\mathrm{E}}$
$\vec{\mathbf{V}}_{\mathbf{T}}$	$\mathbb{R}^{B_T\times B_T}$ set of edges for $\mathbf{T}$
w_{ij}	(i,j)'th entry of $\mathbf{W}$
$\mathbf{W}$	$\mathbb{R}^{N\times N}$ weighting matrix
$\mathbf{W}$	$\mathbb{R}^{N\times\eta}$ component of $\mathbf{A}$
$\boldsymbol{x}$	$\mathbb{R}^{n_{\boldsymbol{x}}}$ state vector
$\boldsymbol{x}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}}}$ state vector at time $\bar{k}$

$\mathbf{x}_i$	$\mathbb{R}^\eta$ point corresponding to the i-th row of $\mathbf{X}$
$\mathbf{x}_\ell$	$\mathbb{R}^\eta$ coordinates lower point defining the squared feasibility region
$\mathbf{x}_\mathrm{r}$	$\mathbb{R}^\eta$ receiver's coordinates
$\mathbf{x}_\mathrm{t}$	$\mathbb{R}^\eta$ transmitter's coordinates
$\mathbf{x}_u$	$\mathbb{R}^\eta$ coordinate upper point defining the squared feasibility region
$\hat{\mathbf{x}}_\mathrm{LS}$	$\mathbb{R}^\eta$ least square estimate of $\mathbf{x}$
$\hat{\mathbf{x}}_\mathrm{MAP}$	$\mathbb{R}^\eta$ maximum a posteriori estimate of $\mathbf{x}$
$\hat{\mathbf{x}}_\mathrm{ML}$	$\mathbb{R}^\eta$ maximum likelihood estimate of $\mathbf{x}$
$\hat{\mathbf{x}}_\mathrm{MMSE}$	$\mathbb{R}^\eta$ minimum mean square error estimate of $\mathbf{x}$
$\mathcal{X}$	state space for a MTT scenario
$\boldsymbol{x}_{1:\bar{k}}$	set of state vectors $\boldsymbol{x}$ from time 1 until time $\bar{k}$
$\boldsymbol{X}_{\bar{k}}$	RFS describing the $L_{\bar{k}}$ targets at $\bar{k}$
$\mathbf{X}$	$\mathbb{R}^{N\times\eta}$ matrix containing the set of Cartesian coordinates of N points in an η dimensional space
$\mathbf{X}_C$	$\mathbb{R}^{N\times\eta}$ coordinate matrix centroids
$\mathbf{X}_{\bar{T}}$	$\mathbb{R}^{N\times\eta}$ matrix $\mathbf{X}$ at time $\bar{T}$
$\bar{\mathbf{X}}_{\bar{T}}$	$\mathbb{R}^{N\times\eta}$ sample mean of $\mathbf{X}_{\bar{T}}$
$\mathbf{y}_{\bar{k}}$	innovation for the observation at $\bar{k}$
$\mathbf{Y}$	$\mathbb{R}^{N\times\eta}$ Gaussian perturbation matrix
z_AOA	AOA observation
z_RSS	RSS observation
z_TOA	TOA observation
$\boldsymbol{z}$	$\mathbb{R}^{n_{\boldsymbol{z}}}$ generic observation vector
$\boldsymbol{z}_{\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{z}}}$ observation at time $\bar{k}$
$\mathcal{Z}$	observation space for a MTT scenario
$\boldsymbol{z}_{1:\bar{k}}$	set of observations from time 1 until time $\bar{k}$
$\widetilde{\boldsymbol{Z}}_{\bar{k}}$	reduced set of observations at time $\bar{k}$
$\boldsymbol{Z}_{\bar{k}}$	RFS describing the $M_{\bar{k}}$ observations at $\bar{k}$
α	parameter monodimensional central and non-central χ distribution
α_RSS	path-loss coefficient
β	parameter monodimensional non-central χ distribution
$\beta^{(j)}$	Boolean variable indicating the emptiness/non-emptiness of the j-th gate
$\beta_{\bar{k}}$	intensities of the RFS $\boldsymbol{B}_{\bar{k}}$
χ^2_n	χ^2 distribution with n degrees of freedom

δ	second parameter Gamma distribution
δ_{ij}	dissimilarity between the i-th and j-th point
$\bar{\Delta}$	reference time
$\mathbf{\Delta}$	$\mathbb{R}^{N\times N}$ generic dissimilarity matrix
$\Delta(\cdot,\cdot)$	subspace distance between the matrices in argument
ϵ	random variable upper bounding ε
$\boldsymbol{\epsilon}_{\bar{k}}^{(ij)}$	$\mathbb{R}^{n_{\boldsymbol{z}}\times 1}$ *innovation* vector associated with the i-th measurement with respect the j-th predicted observation
η	parameter monodimensional central and non-central χ distribution
η	dimension embedding space
γ	first parameter Gamma distribution
γ_{r}	angle difference between the absolute and the relative reference system at the receiver
$\gamma(\cdot)$	intensities of the RFS $\mathbf{\Gamma}_{\bar{k}}$
$\Gamma(\cdot)$	Gamma function
$\mathbf{\Gamma}_{\bar{k}}$	RFS describing the new target birth at $\bar{k}$
κ	first parameter monodimensional Gumbel distribution
$\kappa_{\bar{k}}(\cdot)$	clutter distribution at $\bar{k}$
λ_c	clutter intensity
λ_i	i-th eigenvalue of $\mathbf{\Lambda}$
$\mathbf{\Lambda}$	$\mathbb{R}^{N\times N}$ matrix containing the eigenvalues of $\mathbf{K}$
$\mathbf{\Lambda_a}$	$\mathbb{R}^{N\times N}$ matrix containing the eigenvalues of $\mathbf{K_a}$
$\mathbf{\Lambda}_{\boldsymbol{A}}$	$\mathbb{R}^{B_A\times B_A}$ set of eigenvalues of $\boldsymbol{A}$
μ	mean monodimensional Gaussian distribution
$\bar{\nu}$	average target dynamic
$\nu_{\bar{k}}$	intensity function at $\bar{k}$
$\nu_{\mathrm{S},\bar{k}}$	intensity function for the surviving targets at $\bar{k}$
$\nu_{\Gamma,\bar{k}}$	intensity function for the RFS Γ at $\bar{k}$
ω_{ij}	(i,j)'s entry of $\mathbf{\Omega}$
$\omega_{\bar{k}}^{(j)}$	weight for the j-th component of the GM
$\bar{\omega}_i$	radii i-th Gershgorin disc
Ω	turning rate
$\mathbf{\Omega}$	$\mathbb{R}^{N\times N}$ hollow matrix containing the off-diagonal entries of $\mathbf{K}$
$\mathbf{\Phi}$	$\mathbb{R}^{\eta\times\eta}$ first minor of $\mathbf{\Omega}$
Π_b	permutations of the set $\{1,\dots,b\}$ with $b\in\mathbb{N}\in\{1,2,\dots\}$

$\boldsymbol{\Psi}$	$\mathbb{R}^{N-\eta \times N-\eta}$ lower right block of $\boldsymbol{\Omega}$
ρ	parameter Tikhonov distribution
σ	standard deviation monodimensional Gaussian distribution
σ_d	standard deviation measurement noise on d
σ_y^2	variance single random variables in $\mathbf{Y}$
$\sigma_{\boldsymbol{\Gamma}}$	standard deviation used to model the RFS $\boldsymbol{\Gamma}$
$\sigma_{\boldsymbol{q}}$	standard deviation process noise $\boldsymbol{q}$
$\sigma_{\boldsymbol{r}}$	standard deviation measurement noise $\boldsymbol{r}$
σ_θ	standard deviation measurement angle θ
σ_Ω	standard deviation turning rate Ω
$\boldsymbol{\Sigma}$	$\mathbb{R}^{N \times N}$ singular values of $\mathbf{K}$
$\boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ Cholesky factorization of the state covariance matrix at $\bar{k}$
$\boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}}^{(j)}$	$\mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ Cholesky factorization of the state covariance matrix for the j-th component of a GM
$\boldsymbol{\Sigma}_{\boldsymbol{Q},\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ Cholesky factorization for the process noise covariance matrix $\boldsymbol{Q}$
$\boldsymbol{\Sigma}_{\boldsymbol{R},\bar{k}}$	$\mathbb{R}^{n_{\boldsymbol{z}} \times n_{\boldsymbol{z}}}$ Cholesky factorization for the measurement noise covariance matrix $\boldsymbol{R}$
$\boldsymbol{\Sigma}_{\mathrm{S},\bar{k}}^{(j)}$	$\mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ Cholesky factorization of the state covariance matrix for the j-th surviving component of the GM representing the targets
$\boldsymbol{\Sigma}_{\Gamma,\bar{k}}^{(j)}$	$\mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ Cholesky factorization for the j-th component of the GM modeling the RFS Γ
τ_{t}	transmitter's reference time
τ	second parameter monodimensional Gumbel distribution
$\theta_{i,j}$	angle between the i-th and j-th point as seen from the origin
$\boldsymbol{\Theta}$	$\mathbb{R}^{\eta \times N-\eta}$ second minor of $\boldsymbol{\Omega}$
$\boldsymbol{\Theta}_{\boldsymbol{A}}$	$\mathbb{R}^{B_A \times B_A}$ angle matrix for $\boldsymbol{A}$
$\boldsymbol{\Theta}_C$	$\mathbb{R}^{M \times M}$ angle matrix computed from $\mathbf{X}_C$
$\boldsymbol{\Theta}_{\mathrm{E}}$	$\mathbb{R}^{M \times M}$ angle matrix for $\mathbf{K}_{\mathrm{E}}$
$\boldsymbol{\Theta}_{\mathbf{K}}$	$\mathbb{R}^{N \times N}$ angle matrix for $\mathbf{K}$
$\boldsymbol{\Theta}_{\mathbf{T}}$	$\mathbb{R}^{B_T \times B_T}$ angle matrix of $\mathbf{T}$
υ_i	center i-th Gershgorin disc
υ_i	i-th diagonal element of $\boldsymbol{\Upsilon}$
$\boldsymbol{\Upsilon}$	$\mathbb{R}^{N \times N}$ diagonal matrix containing the diagonal entries of $\mathbf{K}$
$\zeta(\cdot)$	function $f(\cdot)$ or $g(\cdot)$ normalized by $(2\pi)^{\lvert\boldsymbol{x}\rvert/2} \cdot \mathbf{P}_{\boldsymbol{x}}$

ε_θ	Tikhonov distributed angle error
$\boldsymbol{\varepsilon}$	$\mathbb{R}^{N-\eta\times 1}$ first η diagonal elements of $\boldsymbol{\Upsilon}$
$\boldsymbol{\varepsilon}^{(i)}$	i-th cubature point
$\varpi^{(i)}$	i-th cubature weight
$\varphi_{\bar{k}\mid\bar{k}-1}(\cdot\mid\cdot)$	multitarget transition density
$\boldsymbol{\varrho}$	$\mathbb{R}^{\eta\times 1}$ first η diagonal elements of $\boldsymbol{\Upsilon}$
ϑ	angle Jacobi-Givens rotation

ADOA	angle difference of arrival
AOA	angle of arrival
BP	belief propagation
cdf	cumulative density function
CB-MeMBer	cardinality balanced multi-target multi-Bernoulli
Cell-ID	cell identification
CIR	channel impulse response
CKF	cubature Kalman filter
C-MDS	classical multidimensional scaling
CPHD	cardinalized probability hypothesis density
CRLB	Cramér-Rao lower bound
DSSS	direct sequence spread-spectrum
DWT	discrete wavelet transform
E-911	enhanced-911
EDM	Euclidean distance matrix
EKF	extended Kalman filter
FCC	Federal Communication Commission
FIM	Fisher information matrix
FISST	finite set statistic
GM	Gaussian mixture
GNN	global nearest neighbor
GNSS	global navigation satellite system
GPS	global positioning system
HMM	hidden Markov model
ID	identification
IMM	interactive multiple model
IPDA	integrated probability data association

IR	infrared radiation
IT	information technology
JIPDA	joint integrated probability data association
JPDA	joint probability data association
KF	Kalman filter
LBS	location based services
LG	linear Gaussian
LOS	line of sight
LS	least squares
LT	localization and tracking
LM	largest mean
MAP	maximum a-posteriori
MDS	multidimensional scaling
MHT	multi hypothesis tracking
ML	maximum likelihood
MMSE	minimum mean square error
MSE	mean square error
MTT	multitarget tracking
NBP	non-parametric belief propagation
NCV	nearly constant velocity
NLOS	non line of sight
NLS	non-linear least square
NN	nearest neighbor
NNKF	nearest neighbor Kalman filter
OFDM	orthogonal frequency-division multiplexing
OSPA	optimal sub-pattern assignment
pdf	probability density function
pmf	probability mass function
PCA	principal component analysis
PDA	probability data association
PF	particle filter
PHD	probability hypothesis density
PMHT	probabilistic multi hypothesis tracking
POA	phase of arrival
PSD	positive semi-definite

RF	radio frequency
RFID	radio frequency identification
RFS	random finite set
RMS	root mean square
RMSE	root mean square error
RSS	received signal strength
RSSD	received signal strength difference
SDP	semi-definite programming
S-CKF	square-root cubature Kalman filter
SMACOF	scaling by majorizing a complicated function
SMC	sequential Monte Carlo
S-MDS	super multidimensional scaling
SN	strongest neighbor
SNR	signal-to-noise-ratio
SONAR	sound navigation and ranging
SR-LS	squared range least squares
STT	single target tracking
S-UKF	square-root unscented Kalman filter
SVD	singular value decomposition
TDOA	time difference of arrival
TOA	time of arrival
UKF	unscented Kalman filter
UT	unscented transform
UWB	ultra-wideband
WLS	weighted least squares
WSN	wireless sensor network

Contents

1 Introduction to the multitarget localization & tracking problem

1.1 Motivation: why MTT for wireless systems

Localization and tracking is one of the latest major innovations of wireless and personal communications. In the last decade the need for highly personalized information services has thriven the creation of novel location based services (LBS), contributing to make accurate location and tracking information anywhere anytime to become a strategic asset for companies operating in the information technology (IT) business. The core technology behind LBS are positioning and tracking applications.

Although localization and tracking (LT) services were brought to the attention of the mass market through the introduction of global positioning system (GPS) technology, in cellular systems, this revolution started at the end of the 90's, when the US federal communication commission (FCC) mandated the operators to estimate the position of emergency callers with an error of less than 100 meters in the enhanced-911 (E-911) service.

In addition to these more conventional applications, the appearance of new technologies, such as Ultra-WideBand (UWB) with its sub-centimetre accuracy, as well as the improvement and commercialization to the consumer market of well known ones (*e.g.* Wi-Fi), represented the thriving factor that allowed for the development of the new LT services currently available to mass consumers.

In the near future, instead, it can be expected that benefits coming from proper LT solutions could also allow the investigation of techniques able to enhance communications system in terms of performance and efficiency. Indeed, it can be foreseen that location-aware systems will be able to acquire and disseminate information through improved routing, relaying and data fusion techniques based on the LT information. Efficient and reliable localization algorithm will therefore serve the study of strategic cross-layer optimizations for networking and communication functionalities, *e.g.* optimum strategies to perform communications using location awareness.

Although a number of LT algorithms is already available in the literature,

often those methods are designed under specific needs and consequently do not exploit the potential that the different radios currently available offer.

For this reason the design of low-complexity solutions capable to exploit the heterogeneity of information available (ranging, angle, signal strength, hop count etc.) while being still able to cope with dynamic (mobility) of the targets, often more prone to outage (packet loss, activity factor, fading, and so forth), is becoming the biggest challenge in defining future location aware systems.

In addition to the above, the fact that LT services are more and more frequently used in handset devices, such as PDA, mobile phones and so forth, increases the need for novel algorithms that, developed in centralized as well as distributed fashion, can provide the service at a reasonable complexity and possibly without any substantial delay.

In particular, concerning multitarget localization and tracking scenarios, under the assumption of reliable location estimates, it can be expected that the driving factor that will motivate the choice amongst the different solutions will be the computational time employed by the algorithm. With this in mind, the thesis proposes low complexity solutions to a vast variety of multitarget LT scenarios and compares them to the current state-of-the-art.

1.2 Scope and objectives of the thesis

Efficient multitarget tracking algorithms are an important component of multi-user localization and navigation schemes which is attracting much attention in the last few years due to the new frontiers of application in future wireless communications and services [1–3].

Progress has been made on various aspects of the problem including increasing accuracy and robustness of ranging in the presence of non-line-of-sight (NLOS) conditions and interference [4–8], as well as exploiting the geometrical properties of the network topology and ranging statistics in order to combat harsh propagation conditions [9–11]. Due to the diversity of the possible application scenarios of LT systems, *e.g.* wireless sensor network (WSN), maneuvering target tracking, navigation and so forth, it is hard to imagine that a single algorithm will ever be suitable to solve all the possible scenarios.

The scope of this thesis in a broad sense is to develop LT algorithms for multitarget scenarios with particular emphasis on centralized range-based and

range-bearing systems.[1] Starting from the simplest case of *active* systems in which, except for the range measurement, no other side information is assumed, active harsh environments characterized by line-of-sight (LOS) and NLOS conditions in which the measured data are subject to data erasures and biases are investigated.

Heterogeneous observations are also considered in presence of passive MTT in which part of the measurements are due to clutter and no information about the number of targets is provided.

The common denominator of real-time multitarget LT systems is the need for computationally efficient solutions. With this is mind the frameworks and solutions proposed in the Thesis are always aimed at, while ensuring precise and reliable location estimates, the computational complexity of MTT systems as a whole.

1.3 Author's contributions

Original work of the author cover different aspects of multitarget LT systems which was presented in a number of publications, *i.e.* two journal papers [12, 13] and several conference publications [14–22], whose contribution is addressed in detailed in the following.

A non-parametric solution to the active multitarget problem is investigated within the C-MDS framework in dynamic scenarios. Specifically, in [12, 16, 17] it is shown how to statistically characterize the eigenspace associated to the Euclidean kernel matrices used in the C-MDS algorithm. A further contribution of the articles is a statistically-optimized and truncated variation of the Jacobi algorithm, designed specifically to suit the structure of the dynamic MDS kernels used in the C-MDS framework [12, 16, 17]. The modified Jacobi technique proposed, which is founded on the aforementioned statistical eigenspectrum analysis of the associated with the eigen-decomposition of dynamic MDS kernels used in the C-MDS algorithm, has a computational cost which is brought down to a remarkably low complexity[2] of order $\mathcal{O}(\sqrt{N})$ and where N is the size of

[1] Note that many of the solution proposed can be modified to be implemented in distributed fashion.

[2] Measured in terms of the number of plane rotation sweeps normalized to the subspace distance.

the kernel matrix. This, combined with the inherent non-parametricness of the MDS method itself, and non-parametric pre-filtering techniques, such as the one proposed by the author in [14, 15], make a strong case in favor of MDS-based algorithms for future multiuser location-based wireless communication systems.

To compensate for data erasures in the observed distance measurements, the aforementioned method is used together with a Nyström-inspired method [12, 19] which allows, at a limited increase in computational complexity, to exploit all the information available at the LT unit. The same technique proved to be useful as initialization algorithm for optimization techniques where simulation proved that, when used in conjunction with low complex optimization algorithms, *e.g.* the scaling by majorizing a complicated function (SMACOF), it is capable to approach the theoretical limit expressed by the Cramér-Rao lower bound (CRLB) [19]. In the context of multi-source localization scenarios characterized by partial range and angle information and subject to NLOS conditions similarly to [23] in [18] it was shown how to extent the C-MDS framework by rewriting the kernel matrix to describe the edges associated to the graph of the network.

This revised MDS technique is shown to be robust to erroneous angle information. Differently from [23] the proposed solution is shown to cope with data erasures on both the distances between all pairs of points as well as the angles of the triangles formed by all triads of points. In this way the algorithm is made suitable to many multi source localization scenarios of interest, in particular those characterized by networks with meshy topologies of relatively large dimensions, where node-to-node distances and angles can only be measured within vicinities and where the collection of all such information may be a problem. The second main problem of the solution proposed in [23] is that it requires the eigen-decomposition of a kernel matrix whose size increases quadratically with the number of nodes in the network. The author showed in [18] that this problem can be coped with by incorporating the Nyström approximation and the interval analysis tools respectively.

Finally passive MTT targets are covered in [13, 20, 21]. Specifically in [20] it was shown how to incorporate the cubature Kalman filter (CKF) into the probability hypothesis density (PHD) filter implemented as a Gaussian mixture (GM) for multisensor, multitarget scenarios. In [13, 21], due to the Gaussianity assumption in the GM-cardinalized probability hypothesis density (CPHD), we propose to integrate the square-root cubature Kalman filter (S-CKF) into the

GM-CPHD recursion, which is not only the more suitable, *i.e.* the best design under the Gaussian assumption, but also superior to state-of-the-art solutions such as extended and unscented implementation of the Kalman filter (KF), since no inconsistency in the covariance matrix can arise. A novel weighted gating strategy, which exploits the GM implementation of the proposed S-CKF-GM-CPHD filter, is offered to lower the computational time by adaptively increasing the gate sizes in proportion to the likelihood of the single GM components.

1.4 Outline of the thesis

The thesis is structured as follows. Chapter 2 presents a comprehensive overview of the current state-of-the-art for multitarget LT systems in *active* and *passive* scenarios, which serves to contextualize the work presented in thhis manuscript. In Chapter 3 the fundamentals and the algorithms for the different formulations that will be used in the sequel of this thesis to solve the multitarget LT problem are introduced.

Chapter 4 focuses on the STRAIN formulation of the complete *active* MTT problem, showing how to exploit the structure of random Gram matrices emerging in dynamic scenarios within the C-MDS framework. In particular the statistical analysis of Gram matrices using the Gershgorin spectral bounds is performed and a statistically-optimized and truncated variation of the Jacobi technique, which is founded on the aforementioned statistical Gershgorin-theoretical eigen-spectrum analysis and designed specifically to suit the structure of the dynamic MDS kernels under consideration is proposed.

Chapter 5 deals with the data erasure problem and the integration of range-bearing measurement information in active multitarget scenarios. A solution to the first problem, namely the Euclidean distance matrix (EDM) completion is proposed in the form of a kernel approximation solution obtained through the Nyström method. It is also shown that the same technique can be used in conjunction with solutions based on the optimization of a cost function (*e.g.* the STRESS and the SSTRESS cost function). Specifically, within this context it is shown that by coupling low-complexity optimization algorithms with the efficient initialization method proposed is sufficient to bring the performance of low-complexity algorithms such as the SMACOF close to that of far more sophisticated alternatives such as the semi-definite programming (SDP) approach.

The second part of Chapter 5 deals with an algebraic solution that allows for both distance and angle information to be processed algebraically (without iteration) and simultaneously by solving the C-MDS framework over the edges of the graph associated to the network form by the nodes. Due to the flexibility of the method, the algorithm can be executed relying on the Nyström approximation, reducing computational complexity to that of a few matrix multiplications.

Passive MTT scenarios are the subject of Chapter 6. In particular, due to the Gaussian assumption in the GM-CPHD recursion, it is shown how to integrate the S-CKF into the aforementioned filter, leading to the more suitable, *i.e.* the best design and superior to state-of-the-art solutions in which inconsistence in the covariance matrix can arise. Moreover, to lower the computational time of the solution, a novel weighted gating strategy, which exploits the GM implementation of the proposed S-CKF-GM-CPHD filter, by adaptively increasing the gate sizes in proportion to the likelihood of the single GM components is presented. The results reveal that the proposed gating strategy yields considerable savings in processing requirements without suffering any degradation in performance. Final remarks and future works are discussed in Chapter 7.

2 Literature review

In the last 20 years astonishing progresses in the filed of wireless technologies had major influence on a large number of markets and applications. Particularly in the positioning field, supported by mature technologies the initially limited number of applications *e.g.* navigation based on GPS, nowadays extended to global navigation satellite system (GNSS) and emergency calls based on the phase 1 of the FCC-E-911 service, has stimulated an increasing demand for new LT services in a variety of scenarios.

Understandably this diversity of LT applications translates into the need for different system architectures capable of providing services according to the specific requirements imposed by the scenario.

This chapter provides a classification of the building blocks and solutions nowadays available in designing LT systems. Specifically, starting from a characterization of the most common types of information, *i.e.* sensor models, the systems are classified into *active* and *passive* and for each one of those a brief review of the state-of-the-art is provided.

2.1 LT algorithms for wireless systems

A classification of the different LT systems requires the definition of the intended technology as well as the application scenario. A simplified representation of a wireless LT system is offered in Figure 1.

Specifically, the sensor's antenna captures the energy emitted by the targets in the form of radio signals, reflected waves or other forms of waves generating from the targets (*e.g.* infrared radiation (IR)). Together with the desired information, corrupted data, *e.g.* clutter measurements, are also sensed. Therefore, in order to improve the reliability of the measurements, a signal processing block is often added to the sensor unit to remove part of these perturbations on the observations. The last component of the LT system is responsible for fusing the information available, *e.g.* received at multiple sensors, to provide the desired output, *e.g.* target locations in wireless sensor networks or the values of the *state* vector parameter in an LT system.

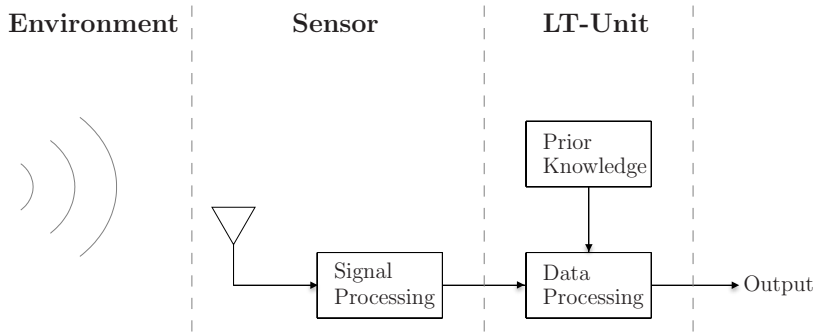


Fig 1. Generic representation of a wireless LT system.

While Figure 1 provides an oversimplified representation of a wireless LT system, a better definition of the functions of the single blocks requires a more precise characterization of the scenario under investigation, as well as the technologies that will be employed in the system. Some of the criteria commonly adopted are:

- *Type of system*: *active* versus *passive*. Active systems assume that during the measurement phase there is an exchange of data between targets and anchor nodes. Differently in passive systems the measurements are *sensed* from the environment in the form of energy that is either reflected or emitted by the target nodes.
- *Sensor model*: any quantity that is directly or indirectly position-dependent, such as time of arrival (TOA), received signal strength (RSS), angle of arrival (AOA), phase of arrival (POA), and that can be measured by the anchor nodes.
- *Architecture*: *centralized* versus *distributed*. Centralized architectures assume that all the information is made available at one node, which is also responsible for recovering the network topology. In distributed systems on the other hand, the location information is recovered by merging the contribution of each node on the basis of the information it gathered from its neighbors.
- *Reference system*: *geographical* versus *relative/symbolic*. In the case of geographical reference systems the solution is a physical location, *i.e.* in the form of latitude and longitude coordinates, differently in presence of relative reference systems the targets positions are expressed with respect to specific sets of points. More generally, relative solutions include *symbolic* systems in

which the output can be in the form of natural-language, *e.g.*, floor and room number.
- *Type of implementation*: *iterative* versus *closed-form*, often related to the computational complexity, namely the scalability of the solution as well as its stability.
- *Processing scheme*: real-time/recursive, batch and fixed-lag solutions. The choice is strongly dependent on the scenario under investigation. For instance, in maneuvering target tracking applications it is desirable to have recursive solution that, at a fixed computational cost, provide continuous updates of the targets locations.
- *A priori* information. Whenever available, prior knowledge, such as dynamic target models or statistical characterization of the measurements, can be used to improve the overall performance.

Obviously the criteria above interrelate amongst themselves. As an example distributed localization algorithms for dynamic scenarios require information that can be gathered in an asynchronous manner between the nodes, *e.g.* TOA measurements, possible knowledge on the kind of motion as well as a common reference system. In the following, we focus on the most common type of information available for LT applications and how they influence the solution.

2.2 Sensor and system models: the input information

Amongst the technologies available to be used in an LT system, a first broad distinction can be made between:

- RF based: WiFi, UWB [7], Bluetooth [24], ZigBee, RFIDs (passive vs active) [25], and so forth
- non-RF based: infrared (*e.g.* "Active Badge" system), ultrasound (*e.g.* "cricket" system) [26], camera vision [27], lasers, and many more.

Once the radio interface is chosen, the next step is to extract from the measurements the observations that are directly or indirectly related to the targets' positions.[3] The most common observations are TOA, AOA and RSS.

TOA observations depend on the distance between a transmitting and a receiving node by means of the propagation velocity c_{p}, *e.g.* the speed of light in

[3]In this work it is assumed that targets are points in space without physical dimensions.

a vacuum. Let $\mathbf{x}_\mathrm{t}$ and $\mathbf{x}_\mathrm{r}$ be the vectors containing transmitter and receiver coordinates, then the relation between the two sensors's coordinates and the measurement is given by [28]

$$z_\mathrm{TOA} = \tau_t + \tfrac{1}{c_\mathrm{p}} \|\mathbf{x}_\mathrm{t} - \mathbf{x}_\mathrm{r}\|_2 + m_n, \tag{1}$$

where τ_t is the reference time at the transmitter and m_n the measurement noise.

Since τ_t needs to be known at both transmitter and receiver, a direct implication of equation (1) is that the devices involved in the measurement phase need to be synchronized amongst themselves. TOA estimates between asynchronous devices are also possible by relying on communication schemes, namely two and three way ranging techniques [29], to infer the value of τ_t during the measurement phase.

A second type of observation commonly encountered in LT systems are AOA, which can be modeled by [28]

$$z_\mathrm{AOA} = -\gamma_\mathrm{r} + \angle(\mathbf{x}_\mathrm{t}, \mathbf{x}_\mathrm{r}) + m_n, \tag{2}$$

where $\angle(\mathbf{x}_\mathrm{t}, \mathbf{x}_\mathrm{r})$ is the angle between transmitter and receiver measured with respect to the geographical reference system, and γ_r is the difference between the absolute and the relative reference system used at the receiver.

The third type of observation commonly used are the ones based on RSS measurements, in which propagation/attenuation models of the transmitted signal are used at the receiver to infer a range estimate. One of such model is [28]

$$z_\mathrm{RSS} = P_\mathrm{t} - L_{d_0} - 10\alpha_\mathrm{RSS} \log_{10} \frac{\|\mathbf{x}_\mathrm{t} - \mathbf{x}_\mathrm{r}\|_2}{d_0} + m_n, \tag{3}$$

where P_t is the power transmitted at $\mathbf{x}_\mathrm{t}$, L_{d_0} the (dB) loss at the reference distance d_0 and α_RSS the attenuation rate of the signal, namely the *path-loss*.

A clear advantage of RSS measurement is that, being already available in every wireless device, RSS-based LT systems become merely a software problem. However, especially in the environments characterized by harsh propagation conditions, *e.g.* indoor scenarios, RSS models are often inadequate to describe the real signal attenuation and multipath effects.

A common feature amongst the types of observation mentioned above is that they all require precise knowledge of a specific parameter, namely τ_t for TOA observations, γ_r for AOA and P_t for RSS observations. When such

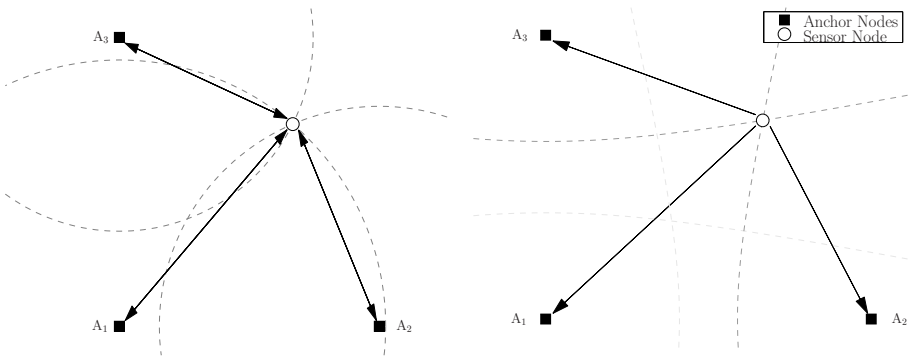
knowledge is not available, the measurement process can be simplified at the expense of performance. In particular, TOA and RSS give rise to time difference of arrival (TDOA) and received signal strength difference (RSSD) when the signal emitted by the target is received at multiple anchor nodes, while AOA becomes angle difference of arrival (ADOA) when the receiver measures the angle difference between two transmitters [28].

The theoretical limits achievable using the measurement models above have been widely investigated in the literature [30, 31]. Following the geometric interpretation of the systems based on the measurement models described above is provided.

2.2.1 Distance information (TOA, TDOA sensors)

Given an η-dimensional embedding space and provided that at least $\eta + 1$ anchor-to-target TOA measurements are available, a geometric solution to the localization problem can be found by *trilateration* [25].

A representation for the 2-D case is provided in Figure 2 from where it is clear that the target location is found by intersecting the circles associated to each one of the anchor-to-target TOA measurement.



(a) Geometry of trilateration algorithm. (b) Geometry of multilateration algorithm.

Fig 2. Representation of a single target system based on TOA and TDOA information.

As mentioned above, to compute TOA estimates using equation (1) anchor-to-target synchronization must be ensured. One convenient way to do so relies on communication schemes [29], which are here indicated by the double direction of the arrows in Figure 2(a).

For the reason above, TOA-based solutions are suitable and often preferred in distributed scenarios or mash networks where the communication between agents is allowed. Alternatively, if the targets are mere transmitters and the anchor nodes are connected and synchronized amongst themselves, *e.g.* cellular network scenarios where a backbone infrastructure can be used to share a common reference time [25], then TDOA-based algorithms, usually referred to as *multilateration* [25] techniques, can and should be used. As Figure 2(b) shows, from a geometrical perspective multilateration algorithms find the intersection point of the hyperbolas whose focal points are the pair of anchors used to compute the TDOA measurements.

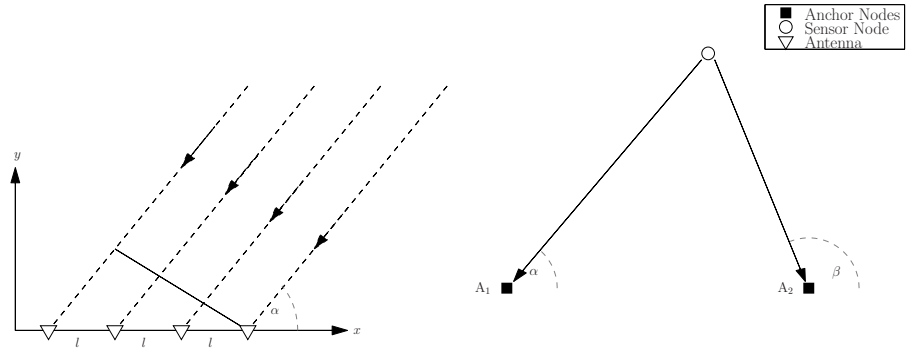
Depending on the synchronization requirement, trilateration and/or multilateration schemes can be used with *e.g.* UWB radios, direct sequence spread-spectrum (DSSS), Orthogonal frequency-division multiplexing (OFDM) systems, POA-based system, lasers radio doppler shifts, ultrasound signal, and many other technologies.

2.2.2 Angle information (AOA sensors)

AOA-based systems find the target location by intersecting the direction lines obtained from the angles measured at the receivers by means of sensor arrays [32]. The algorithms used to find the target location on the basis of AOA observations are often referred to as *triangulation* techniques.

A geometric representation of the problem in a 2-D scenario is provided in Figure 3(a). Notice that AOA measures are computed according to the sensor array orientation, from which it follows that the correction terms γ_{r} at the anchors must be considered during the triangulation phase.

Although AOA-based systems can and are commonly used in radar scenarios [27], the application of AOA-based LT systems in wireless networks is quite limited.



(a) Sensor array for AOA estimation. (b) Geometry of triangulation algorithm.

Fig 3. Representation of a single target system based on AOA information.

Some of the reasons are: the relatively large and expensive hardware, namely the sensor array represented in Figure 3(a), the quality of the angle measurements that can be particularly compromised in presence of shadowing and multipath[4] as well as the need to compensate γ_r amongst all the anchor nodes. Therefore, although the application of AOA-based algorithms can be still envisioned in cellular network scenarios characterized by a considerable number of anchor nodes with *limited* connectivity, due to the problems mentioned above they are not practical in distributed systems where there is no clear distinction between anchor and target nodes and where all agents are allowed to move.

2.2.3 Connectivity information (proximity sensors)

Although equation (3) shows that RSS measurements provide range estimates and therefore can be used in conjunction with the trilateration and multilateration schemes described in Section 2.2.1, it is also possible to use simplified RSS models, *e.g.* detection models, to use localization schemes based on mere connectivity information. Due to the intrinsically low cost of the information used, and more importantly because of the wide number of emerging applications requiring less stringent location notions [33], systems based on this type of information gained significant interest. As Figure 4 shows, using connectivity information only it is

[4]Also the accuracy of AOA-system degrades as the source move farther from the measuring unit.

possible to define a *feasibility region* in which the target is ensured to be.

Moreover, although a punctual estimate of the target estimate can be inferred directly once this feasibility region is known, *e.g.* its barycenter, more sophisticated LT techniques integrate this notion about the feasibility region as a constraint into optimization algorithms that are capable of improving the estimates for the targets' locations [33].

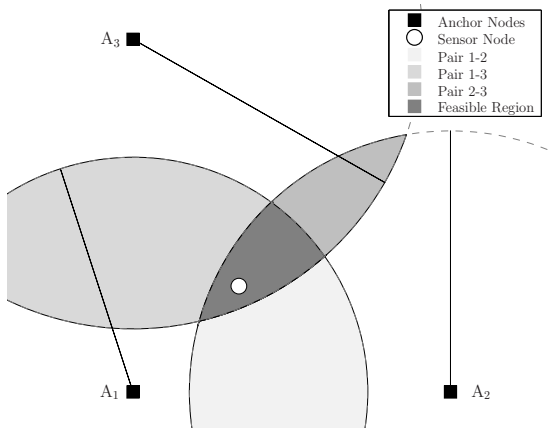


Fig 4. Representation of a single target system based on connectivity information.

This is particularly true for TOA solutions where the feasibility region is used to compensate eventual bias on the range measurements [4, 34, 35].

In addition to RSS-based observations, proximity based LT systems can also rely on radio frequency identification (RFID) and cell identification (Cell-ID) information. However one problem of this approach is represented by the quality of the information which is strongly dependent on the density of the anchor nodes as well as the detection model used.

2.2.4 Other types of information

Other types of information, *e.g.* root mean square (RMS) delay spread, channel impulse response (CIR) or maximum excess delay, can be used in pattern-matching systems, also known as fingerprinting, to provide LT services. This type of systems consist of two phases. Namely, an *off-line training phase*, during

which a representation of the environment is built on the basis of a measured model and the sensed information, and an *on-line* phase during which the target's location is estimated by matching the measured information with the mapping of the environment built during the off-line phase.

Figure 5 provides a representation of a fingerprint system for navigation purposes, *e.g.* the system proposed in [26], in which the computation is performed at the target node. Although the approach can be successfully employed in specific applications, in general the time used during the off-line phase as well as the need for frequent database updates to account for changes in the environment, represent strong limitations often preventing its usage in practical scenarios.

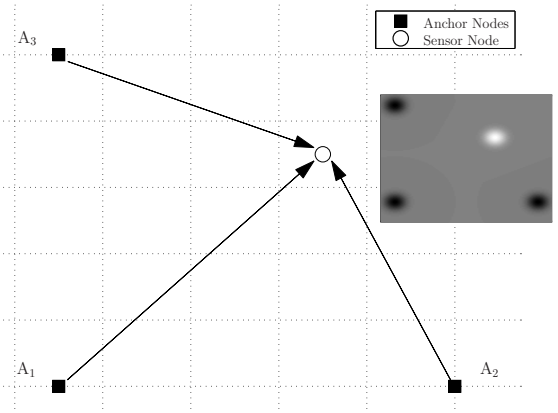


Fig 5. Representation of a fingerprint system for navigation.

2.3 Active systems

As already anticipated in Section 2.1, in this thesis, *active systems* are LT systems in which target nodes participate in the measurement phase by exchanging data with the anchor nodes. This, in turn, implies that:

- targets are distinguished by an identification (ID) number
- observations are uniquely associated to the targets from which they generated, namely there is no *data association problem*
- depending on the type of architecture, both targets and anchor nodes can be transceiver (*e.g.* in cooperative TOA-based scenarios) or targets are

transmitters and the anchors are receivers (*e.g.*, TDOA-based scenarios).

Concerning the solutions that can be used in active LT problems, a first distinction must be made between *parameter* and *state* estimation approaches. The former refers to the estimate of time-invariant quantities, namely multitarget localization problems, and the latter to the estimate of the states[5] of a time-variant system, commonly encountered in tracking scenarios.

A brief classification for the state-of-the-art solutions to the active LT problem is provided below, in which a distinction is observed between non-Bayesian and the Bayesian techniques.

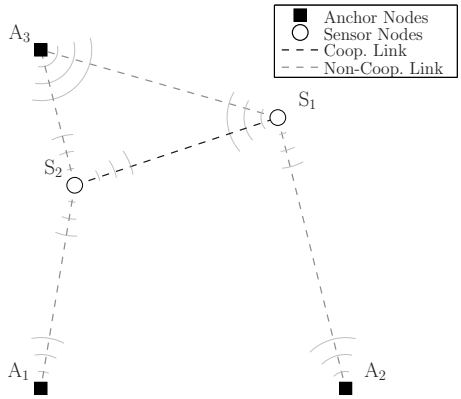


Fig 6. Representation of a cooperative multitarget scenario.

2.3.1 Non-Bayesian solutions

In this category belong all those methods that can be used to solve the standard *Euclidean embedding* problem, namely: given a dissimilarity [36, 37] matrix $\mathbf{\Delta}$, find the configuration of points in an Euclidean embedding space [38, 39] such that the mutual distances amongst the points match the corresponding dissimilarities in $\mathbf{\Delta}$.

Depending on the kind of formulation used in the algorithms, the solutions of

[5] As explained later, states are the variables of interest whose value must be coputed by the algorithm, *e.g.* coordinates, velocity and so forth.

the aforementioned problem can be distinguished between indirect and direct methods

1. Indirect parameter estimation methods

As suggested by the name, indirect methods find a solution which is indirectly related to the targets' locations,*e.g.* in form of a distance or a kernel matrix. Once the solution space, *e.g.* the EDM cone is defined[6], these methods, also called *matrix proximity* methods [40, 41], find the closest matrix to $\mathbf{\Delta}$ with respect to a specified norm, *e.g.* the Frobenius norm. More importantly they often rely on a *convex* formulation of problem which can be solved optimally [40, 42].

Amongst such methods, a solution originally used in psychology as a data analysis technique [38] and that is becoming increasingly popular in the field of localization [12, 43–45], is the C-MDS algorithm [37]. The C-MDS is a particular formulation of the proximity problem on the space of positive semi-definite (PSD) matrices [40] and it is appealing since it represents the optimal closed form solution to the problem with respect to its cost function [46, 47].

However, since the method relies on the eigen-decomposition of the typically large Euclidean kernels constructed from the observations, its usage in *dynamic* scenarios where continuous updates of the solution are required could be problematic [12]. Moreover, in its standard formulation the C-MDS approach does not cope with possible data erasures in $\mathbf{\Delta}$, which is typical in localization scenarios characterized by mesh topologies.

More complex approaches relax the matrix approximation problem to a weighted, nearest, completion problem over the space of PSD matrices, often solved using SDP techniques [42, 48–50]. Differently from the C-MDS, the SDP weights the single dissimilarities in $\mathbf{\Delta}$, solving the data erasures problem as well as the problem of considering the dissimilarities in proportion to the quality of their estimates. In [51, 52] an SDP formulation over the space of Euclidean distance matrices was proposed. While the formulation of the problem is somehow more involved than in [50] the output of the method is the complete EDM that approximates the entries in $\mathbf{\Delta}$. This implies that the SDP algorithm is only a block of the localization system since an additional algorithm, *e.g.*

[6]In the subspace of symmetric matrices, the closed convex set of all Euclidean distance matrices is what is generally referred to as the EDM cone.

the C-MDS discussed above, is still required to compute the target estimates. Extensions of the SDP formulation to TDOA scenarios were proposed in [53, 54].

Although SDP formulations allow to pose the problem as a convex optimization problem [42, 55], they doi it at the price of a high computational complexity. Some recent solutions that try to relax the computational requirements of the aforementioned SDP formulations were recently proposed in [56, 57].

2. Direct parameter estimation methods

Differently from the above direct non-Bayesian methods[7], solve the localization problem directly but without ensuring its convexity [58]. Indeed, given a set of data, these type of solutions adjust the model parameter values such that the distance between an objective function and the data is minimal. This is done applying numerical optimization algorithms to the cost function used to represent the problem.[8] In particular, given the observation vector $\mathbf{z} \in \mathbb{R}^{n_\mathbf{z}}$ and $f(\cdot)$ as the function relating measurements and the vector parameter $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}}$, most of the formulations used to do parameter estimation rely on either of the following:

– minimization of a least squares (LS) function

$$\hat{\mathbf{x}}_{\mathrm{LS}} \triangleq \arg\min_{\mathbf{x}} \|\mathbf{z} - f(\mathbf{x})\|^2 \tag{4}$$

– maximization of a likelihood function

$$\hat{\mathbf{x}}_{\mathrm{ML}} \triangleq \arg\max_{\mathbf{x}} p(\mathbf{z}|\mathbf{x}). \tag{5}$$

While the LS approach does not assume any characterization of the noise statistic affecting the observations, as equation (5) shows, the maximum likelihood (ML) approach requires the conditional distribution $p(\mathbf{z}|\mathbf{x})$. However, under the condition of zero mean Gaussian noise the ML and the LS estimators are proved to be equivalent.[9]

[7] These methods, also known as *Fisher* methods have a lower bound on the error covariance matrix which is given by the inverse of the Fisher information matrix (FIM), namely the CRLB [30].

[8] Often the cost function is *non-convex* and therefore there is no guarantee that the solution corresponds to the global minima.

[9] In the more general case of multiple observations affected by zero mean Gaussian noise with different standard deviations the ML is equivalent to a weighted least squares (WLS) estimator.

Since the ML estimation applied to sensor localization scenarios is generally a *non-linear* and non-convex optimization problem whose global solution is hardly obtained, it is often preferred to rely on LS formulations that can solve the localization problem without requiring a full characterization of the measurement statistics. In the attempt of balancing between the sub-optimality of the solution and its complexity many different LS algorithms have been proposed in the literature [58]. A first broad distinction can be made between *gradient search* and *direct search* methods.

Gradient-based methods depend on the derivative of the objective function and therefore can only be applied to differentiable problems[10]. Steepest descent methods use the gradient to set the direction of convergence. Newton methods usually have faster convergence than the steepest descent methods but require at each step of the optimization the Hessian matrix, namely the inverse of the second derivative of the objective function[11][58]. The Levenberg-Marquardt method combines the two approaches above using a PSD approximation of the Hessian matrix whenever it gives good convergence results and the steepest descendent otherwise [59].

Although suboptimal, gradient-based methods are used in localization applications in both centralized as well as distributed localization scenarios [60–64].

Modifications of the LS objective into squared range least squares (SR-LS) are also possible and were discussed in [65, 66] in the context of source localization for TOA and TDOA schemes, where it was also shown that the SR-LS approach is only sub-optimal compared to the LS method. More importantly, some algorithms modify the LS objective function during the optimization phase to improve the speed and the properties of the convergence of the algorithm. One such example is the SMACOF algorithm [67, 68], where instead of the standard LS objective function, its majorized version is used at each step of the optimization, resulting in a monotone and fast convergence method [69, 70].

Direct search methods aim at finding the global minimum of the objective function and although they do not rely on derivatives they still require that the objective function is continuous [71]. Grid search, random walk (which is a

[10] Although the algorithms would require the analytical expression of the gradient, numerical approximations are also possible [58].
[11] Quasi-Newton methods avoid to recompute the Hessian at each step.

variation of the random search approach), multi-start and tunneling methods are few of the solutions available in the literature [38, 72]. However one common problem of direct search methods is that they are usually too computationally demanding for real time applications.

Other approaches that modify the LS cost function to improve the convergence of the algorithm exist, some recent solutions were proposed in [73, 74].

2.3.2 Bayesian solutions

Let the parameter $\boldsymbol{x}$ be a random variable with a prior probability density function (pdf) $p(\boldsymbol{x})$ and $\boldsymbol{z}_{1:\bar{k}}$ the set containing the observations from time 1 until time $\bar{k}$, namely $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{\bar{k}}\}$. Then the inference process on $\boldsymbol{x}$ operated by the Bayesian theory can be understood as the *correction of the prior subjective probability distribution by the objective data* $\boldsymbol{z}_{1:\bar{k}}$ [75], which is based on the following building blocks:

- *prior model* $p(\boldsymbol{x})$: representing the subjective belief for the parameter $\boldsymbol{x}$ before the realized measurement value is considered,
- *likelihood model* $p(\boldsymbol{z}|\boldsymbol{x})$: defining the stochastic relationship between $\boldsymbol{z}$ and $\boldsymbol{x}$.

Under the assumption that a realization of $p(\boldsymbol{x})$ occurred and that its value is constant during the measurement process, then the solution of Bayesian estimators is represented by the *posterior* probability distribution conditioned on the data $p_{\bar{k}}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}})$, which is computed using Bayes' rule [30] as

$$p_{\bar{k}}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}}) = \frac{p_{\bar{k}}(\boldsymbol{z}_{1:\bar{k}}|\boldsymbol{x})\, p_{\bar{k}-1}(\boldsymbol{x})}{p_{\bar{k}}(\boldsymbol{z}_{1:\bar{k}})}, \tag{6}$$

where

$$p_{\bar{k}}(\boldsymbol{z}_{1:\bar{k}}) = \int_{\mathbb{R}^{n_{\boldsymbol{x}}}} p_{\bar{k}}(\boldsymbol{z}_{1:\bar{k}}|\boldsymbol{x})\, p_{\bar{k}-1}(\boldsymbol{x})\, d\boldsymbol{x}. \tag{7}$$

Once $p_{\bar{k}}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}})$ is available, the estimate of $\boldsymbol{x}$ based on $p_{\bar{k}}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}})$ becomes a decision theory problem which is solved by performing expectations and maximizations (or equivalently minimizations). Amongst all possibilities, two choices commonly employed to estimate $\boldsymbol{x}$ from $p_{\bar{k}}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}})$ are the minimum mean square error (MMSE)

$$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \mathrm{E}[\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}}] \triangleq \int_{\mathbb{R}^{n_{\boldsymbol{x}}}} \boldsymbol{x} \cdot p_{\bar{k}}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}})\, d\boldsymbol{x}, \tag{8}$$

and the maximum a posteriori (MAP) estimator[12]

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}} \triangleq \max_{\boldsymbol{x}} p_{\bar{k}}\left(\boldsymbol{x}|\mathbf{z}_{1:\bar{k}}\right). \tag{9}$$

Depending on the problems at hand the Bayesian solutions can be distinguished in:

– parameter estimation techniques for time invariant system
– *state* estimation techniques for time variant systems.

In the context of LT applications parameter estimation solutions are commonly employed in static scenarios where the single target locations are inferred on the basis of anchor-to-target as well as target-to-target measurements[13]. For instance in [76] it was proposed to represent the cooperative multitarget case by a graph whose solution can be found using non-parametric belief propagation (NBP) algorithms, namely a distributed Bayesian approach.

Differently from the above, when targets are assumed to evolve in time according to a stochastic model[14], a *state-vector* is used to summarize the past of the system in a probabilistic sense [77]. Starting from the KF which was originally proposed in [78] to deal with linear Gaussian (LG) systems, over the years a huge number of approximations have been proposed to solve the nonlinear single target tracking (STT) filtering problem that arise in real systems[15].

Following [77] a broad classifications of the Bayesian solutions proposed to solve the filtering problem can be made between:

1. Analytic approximations
2. Numerical approximations and Gaussian sum filters

To the first group belong all those solutions that solve the problem by linearization, such as the extended Kalman filter (EKF). Due to their good performance and relatively low complexity these solutions are still applied to a large number of real LT applications [27, 30]. Solutions based on numerical

[12] Notice that the ML solution can be interpreted as the output of a MAP estimator with uniform a priori distribution for $\boldsymbol{x}$.
[13] For multiple observations the estimates for $\boldsymbol{x}$ can be *recursively* updated on the basis of the new observations [30]
[14] Usually the dynamic models are restricted to be probabilistic Markov sequences [30].
[15] Approximations are necessary to make the computation involved in the Bayesian framework feasible.

approximations involve numerical integrations to solve the multidimensional integrals appearing in the Bayesian recursions, while GM solutions approximate $p(\boldsymbol{x}|\boldsymbol{z})$ by a weighted sum of Gaussian functions [79].

Although GM solutions were extensively applied to maneuvering target tracking problems, novel techniques aiming at improving the weights and moments computations are still subject of intensive research [80–82].

3. Sampling approaches
4. Particle filters.

Sampling approaches approximate the posterior density $p(\boldsymbol{x}|\boldsymbol{z})$ by a defined function, *e.g*, the Gaussian function, whose moments are then captured by a small set of deterministically chosen points. The unscented Kalman filter (UKF) and the CKF are successful examples of these type of filters that have been intensively applied to tracking scenarios [83, 83–85].

The last suboptimal solution to the nonlinear filtering problem is based on the particle filter (PF) method which perform sequential Monte Carlo (SMC) estimation using point mass representation of the posterior probability density [86]. Although the approach is well known since several decades, only recently it gained popularity in nonlinear filter applications, such as target tracking problems [77, 87, 88] due to the inclusion of the *resampling* step [89].

While the state estimation techniques mentioned above are mainly concerned with STT scenarios, recently the implementations of distributed KFs [90–92] and, more general distributed Bayesian estimators for statistical inference extended the application of the Bayesian paradigm to dynamic cooperative scenarios [93–95]. Details about the implementation and the theory behind those algorithms are offered in [96, 97].

2.4 Passive systems

Unlike active systems, in a passive setting targets have no direct involvement in the measurement phase Consequently the output of the signal processing block in Figure 7 are observations related to energy either transmitted or reflected by the targets.

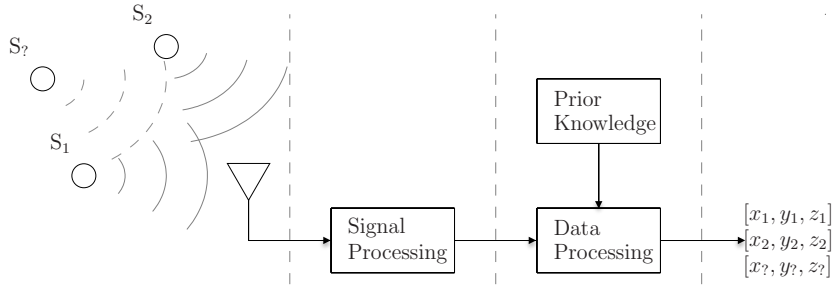


Fig 7. Representation of a passive system.

An example is the sound navigation and ranging (SONAR) system which can be used to listen for either the sound originated by other vessels or the eventual echoes generated from other vessels in response to pulses of sound emitted by the LT system itself [98–100].[16]

While the solutions discussed in Section 2.3 could be used in STT scenarios, in the MTT case, the interdependence between target's locations and measurements requires more sophisticated solutions. The uncertainty on the source of the measured data is what is generally referred to as the *data association* problem and it arises in the more general case of an unknown number or targets and/or when sensors are subject to *false alarm* and *clutter measurement*.

Therefore a first distinction can be made between passive STT settings, in which the data association ambiguity is a consequence of clutter and false alarm measurements only, and the passive MTT settings in which there is a further degree of uncertainty due to the presence of an unknown number targets, possibly more than one.

[16]Strictly speaking the two SONAR architectures are usually referred to as *active* and *passive* systems. This should not be confused with the more general definition of active and passive system given in Section 2.3-2.4 which is used in this work.

Generally speaking, MTT solutions are a generalization of STT approaches and aim at providing consistent estimates of the number of targets in the scenario as well as their trajectories.

A brief review of the conceptual solutions of the MTT problem [27, 101, 102] that rely on explicit associations between measurements and tentative tracks is offered in Section 2.4.1. A recent formulation of the problem based on the finite set statistic (FISST) framework [103], which does not necessitate any association step, is introduced in Section 2.4.2.

2.4.1 Legacy methods

Nearest neighbor and track-splitting methods

The nearest neighbor (NN) solution is the simplest and a relatively low complexity method that solves the data association problem. As suggested by the name, it is based on the assumption that measurements[17] are generated from the closest, in a likelihood sense, point in the space. While this approach is only suitable to solve the STT problem [102] extensions to the MTT scenario exist under the name of global nearest neighbor (GNN) and are based on the global optimization of a cost function, *e.g.* the sum of the likelihood measures. Also know as 2-D assignment, this was the first algorithm designed to solve the MTT problem. A first enhancement aimed at improving the tracking performance consisted in combining the NN method in STT settings or the GNN solution in MTT scenarios together with Bayesian filters. An example of this is the nearest neighbor Kalman filter (NNKF) solution where in the update phase of the recursion the measurement closest to the prediction resulting from a KF is utilized to estimate the target state vector. A variation of the NNKF is the strongest neighbor (SN) filter proposed in [104]. However, the impact of fault assignments can be particularly problematic when dealing with filters since it could make them diverge from the true track. To improve the target-to-measurements assignment more sophisticated solutions were proposed [101].

For instance, under the principle that better assignments can be expected introducing a lag in the computations and using multiple sets of measurements at different times, new approaches based on the construction of decision trees

[17] Only the current set of information is used to make the assignment.

describing the likelihood of different associations at different times were proposed in [102].[18] Due to the combinatorial nature of the methods, these *track-tree* approaches suffer from an explosion in the number of branches used to represent the possible outcomes for the trajectories.

In response to this problem a measure of track likelihood was included in the approach to serve as basis for different pruning schemes, such as:

– delete unlikely tracks
– keep a fixed number of tracks
– merge track nodes

with the latter pruning scheme suggested in order to remove the unrealistic condition occurring in the track-splitting approach that tracks may share measurements. A solution to this problem was offered in the form of a *batch* solution in [105] with the *joint likelihood method*.

However a common problem for all the methods mentioned above is that they do not handle track initialization and termination.

Multi-hypothesis tracking methods

The multi hypothesis tracking (MHT) filters are an improvement to the MTT of the solutions discussed in the previous section. At the core of the MHT filter is the concept of *hypothesis*, which is defined as the exhaustive association of all measurements to all possible disjoint tracks.

The filter starts by building a tree whose leaves are the sets of hypotheses containing active tracks and operates by recursively updating the posterior probability of each hypothesis on the basis of the current observations [27]. At each iteration only the hypotheses with higher probability are selected, making the MHT filter essentially a MAP estimator. Once the most likely hypotheses are selected the target's states are updated using different filters such as standard KFs or interactive multiple model (IMM) methods[19].

Differently from the NN solutions, the problem of track initialization and termination is inherently handled by the filter during the update phase when

[18] These solutions consider as *tracks* the *sequence of measurements originating from the same point in the space* [102].
[19] The MHT is intrinsically an MTT filter although it is possible to modify it to work in the STT scenario.

new measurements are associated to either current tracks, new tracks or clutter measurements.

Depending on the strategy used to create the hypotheses the MHT implementations can be classified into *measurement-oriented* and *track oriented* [27]. In the former solution, which was also the first to be proposed in [106] for STT scenarios, new hypotheses are retained and propagated in time. In turn in the track-oriented approach only the tracks are propagated in time and the hypotheses are reproduced each time from the tracks. This formulation is often preferred over the original measurement-oriented approach since less computationally demanding.

To obviate the exponential complexity inherent in the approach, the MHT filter is used together with pruning schemes that limit the branching in the hypothesis tree by clustering the tracks, merging the track histories or pruning the low-probability branches from hypothesis trees [107]. Although in its standard formulation the MHT filter makes hard associations/decisions, a novel variation of the filter which bases the measurements to targets association on soft decisions exists under the name of probabilistic multi hypothesis tracking (PMHT) [108] which admits an efficient implementation under the name of turbo PMHT [109] .

Probability data association methods

Despite the pruning stage, the MHT filter still suffers from an exponentially growing complexity. This motivated the investigation of new data association schemes that, albeit suboptimal, could be implemented with almost fixed computational complexity. One such solution is the probability data association (PDA) filter in which, differently from the NN approach where only the closest measurement is used for the update stage, each single measurement is considered in proportion to the probability that it was generated by a target [110].

In the PDA the update is performed using the average of all measurements weighted by the corresponding posterior probability computed through the Bayesian recursion. Although the PDA filter was originally proposed to cope with STT scenarios in which the target track is already initialized, the joint probability data association (JPDA) extended the approach to MTT applications with *fixed* and *known* number of targets.

The recursion used in the JPDA filter is the same used in the PDA, with

the exception that instead of *normal* association probabilities *joint* association probabilities are computed to avoid conflicting measurement to track assignments. However, the computation of the joint association probabilities grows exponentially with the number of targets and measurements. In addition to this complexity problem, the JPDA filters operate under the assumption that a *known* and *fixed* number of targets are allowed in the scenario, meaning that the filter cannot cope with track initiation and deletion.

To overcome this problem a generalization of the PDA allowing target appearance and disappearance was proposed in [111] under the name of integrated probability data association (IPDA) and in [27, 112] as the joint integrated probability data association (JIPDA) filter.

2.4.2 State-of-the-art method

A common feature of the methods discussed in Section 2.4.1 is that current targets' states are inferred after fitting the current set of measurements to the existing tracks. Those solutions are therefore often referred to as *report-to-track* methods [113]. They solve the MTT problem relying on measurement to track associations and have the advantage that the target identities are a direct outcome of the algorithms. However, when wrong associations occur, estimation errors may result.

A recent formulation which uses the FISST theory developed by Mahler [103, 114] to cast the MTT problem into the random finite set (RFS) framework is of particular interest since it requires no assumption on the number of targets and does not rely on the idea of hypothesizing associations between measurements and targets. However, since the approach avoids solving the data association problem, it only provides the target's locations without identities.[20]

A brief overview on the RFS principle, the formulation of the MTT problem based on this framework and the conceptual solutions that it offers is presented in the following.

[20] Labeling-algorithms can be used in conjunction with this solution to recover this information [115].

The random finite set framework

The MTT methods discussed in Section 2.4.1 can be seen as the outcome of a series of approximations, such as the one in the *correction-update* stages of the Bayesian recursion used to update the tentative tracks or the ones made to include target initialization and termination. A different perspective is offered by Mahler, who, in a series of works suggested to consider the tracking problem as a whole first, offered its conceptual solution in the form of a generalization of the Bayesian solution to the MTT problem and finally proposed practical implementation of this based on an approximation of the entire model [113, 114, 116].

Using the FISST tool as a basis, in particular the notion of *set derivatives* and *set integrals*, Mahler built a calculus of probability for RFSs making therefore available to the field of RFS ideas from statistic (*e.g.* ML estimate) as well as information theory (*e.g.* Kullback-Leibler divergence) [103]. Then using the generalization of likelihood and prior for RFS Mahler constructed the conceptual solution to problem in the form of the Bayesian recursion for the MTT filtering problem allowing clutter and false alarm measurements as well as target appearances and disappearances directly into the formulation.

The probability hypothesis density approximation

As discussed in Section 2.3.2 for the active STT problems, the multivariate integrals involved in the Bayesian recursion rarely have closed form solution [30] and become even more difficult to solve once the problem is generalized to the passive MTT scenarios by means of the RFS framework.[21]

Once again Mahler, resembling the success of moment matching methods often used in STT problem, *e.g.* the KF-based solutions, proposed to propagate in time only the first moment of the entire MTT posterior pdf, also known as the PHD function.

One major difference between the the entire posterior pdf and the PHD functions is that while the former is defined over the state space of the RFS, the PHD is defined over the space of the single state vector. Consequently, the PHD is not only less complex but also fixed dimensions in time. In addition when the

[21] Under such circumstances the dimension of the state vector describing all targets' parameters become an unknown random number.

PHD is integrated over a portion of the state space it returns the number of targets in it, suggesting that the target's locations can be found by the local maxima of the function.[22]

Since RFSs are fully characterized by a pdf and a cardinality number, with the CPHD Mahler improved the approximation for the MTT posterior pdf by propagating in time both the PHD function and the probability mass function (pmf) for the discrete random variable describing the targets' cardinality.

Although both the PHD and the CPHD recursions represent a substantial simplification compared to the propagation in time of the entire MTT posterior density, they still require the computation of multidimensional integrals. Due to this reason practical implementations of the PHD and the CPHD filters have been proposed in the form of SMC methods [117] and GMs [118, 119]. Both implementations have been shown to converge to the true PHD function as the number of particles and Gaussian Mixture components increase [120] and they both compared favorably when compared to JPDA and MHT solutions [121].

2.5 Summary and discussions

After the broad overview offered in this chapter it should be clear that a unique solution to the variety of possible LT multitarget scenarios is not plausible. Depending on the criteria covered in Section 2.1, different mathematical models/formulations are required to formalize the problem in the best possible way. In particular Chapter 3 introduces the formulations and the algorithms that will be employed to solve the different LT multitarget scenarios covered in the thesis.

Moreover, although different solutions might achieve the same performance in terms of accuracy, another important aspect that must be considered while dealing with multitarget problems is the *scalability* of the solution, which in turn is affected by the computational time required by the method under investigation. This is one of the driving aspects that motivated the work proposed in the sequel of this thesis.

Specifically, starting from the *active*-complete case, which is the subject of Chapter 4, the proposed framework is extended to the case of incomplete and

[22]Since the PHD does not integrate to one it is not a pdf.

heterogeneous information in Chapter 5.

Chapter 6 instead uses the emerging RFS formulation of the *passive* MTT problem mentioned in Section 2.4.2 to cast and solve the problem into a Bayesian framework. Even in the case of *passive* scenarios, incomplete range-based and incomplete range-bearing systems are considered modifying the system model used in the algorithms.

3 Fundamentals of multitarget localization and tracking

This chapter introduces the formulations used to solve the LT multitarget scenarios discussed in this thesis.

Starting from a non-parametric approach based on a matrix-proximity formalization of the problem, more complex solutions based on LS and Bayesian formulation are discussed as possible approaches to deal with active LT scenarios.

The conceptual solution to the passive MTT scenario relying on the theory of FISST and a RFS formulation of the problem is introduced in the second part of the chapter. For each formulation a short description of the algorithms used later in the thesis is also provided.

3.1 The multitarget LT problem with active sensors

Let $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ denote the matrix containing the set of Cartesian coordinates for N points in an η-dimensional space. Then, the Euclidean distance between two generic points $\mathbf{x}_i$ and $\mathbf{x}_j$ is

$$d_{ij} = d_{ji} \triangleq \left(\sum_{k=1}^{\eta}(\mathbf{x}_{ik} - \mathbf{x}_{jk})^2\right)^{\frac{1}{2}}. \tag{10}$$

Also let the squared *Euclidean Distance Matrix* EDM $\mathbf{D}$ be such that $[\mathbf{D}]_{ij} = d_{ij}^2,\ \forall(i,j) = (\{1,\cdots,N\},\{1,\cdots,N\})$, which in matrix form yields

$$\mathcal{D}(\mathbf{X}) = \mathbf{D} = \begin{bmatrix} 0 & d_{1,2}^2 & d_{1,3}^2 & \cdots & d_{1,N}^2 \\ d_{2,1}^2 & 0 & d_{2,3}^2 & \cdots & d_{2,N}^2 \\ d_{3,1}^2 & d_{3,2}^2 & 0 & \cdots & d_{3,N}^2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ d_{N,1}^2 & d_{N,2}^2 & d_{N,3}^2 & \cdots & 0 \end{bmatrix}. \tag{11}$$

While a detailed description of the properties related to $\mathbf{D}$ is available in [40, 41, 48] it is important to realized that since $\mathbf{D}$ is a symmetric matrix with zero diagonal elements, only the $M = \binom{N}{2} = N(N-1)/2$ entries of its upper/lower triangular part are informative.

Using equation (10), the distance d_{ij}^2 can be written as

$$d_{ij}^2 = \mathbf{x}_i \cdot \mathbf{x}_i^{\mathrm{T}} + \mathbf{x}_j \cdot \mathbf{x}_j^{\mathrm{T}} - 2\mathbf{x}_i \cdot \mathbf{x}_j^{\mathrm{T}}, \tag{12}$$

which is the well known *cosine-law* [40].

It follows that the generic ij-th entry of $\mathcal{D}$ is related to the triangle obtained from the points $\mathbf{x}_i$, $\mathbf{x}_j$ and the origin of the reference system used. Moreover, since any rigid transformation of $\mathbf{X}$ produces the same EDM $\mathbf{D}$ [40], algorithms that rely exclusively on $\mathbf{D}$ to infer $\mathbf{X}$ can only find the solution up to an *isometry* in the embedding space. To resolve this uncertainty it is necessary to perform a rigid transformation of the solutions which can be done by the Procrustes [37] transformation detailed in Appendix 1.

3.1.1 The STRAIN formulation

Let $\mathbb{D}_N(\eta)$ be the space of $N \times N$ distance matrices generated by N points in an η dimensional space and let δ_{ij} be an arbitrary dissimilarity between $\mathbf{x}_i$ and $\mathbf{x}_j$. Now consider the squared dissimilarity matrix $\mathbf{\Delta} \in \mathbb{R}^{[N \times N]}$ such that $[\mathbf{\Delta}]_{ij} = \delta_{ij}^2$ and its transformation

$$\mathcal{K}(\mathbf{\Delta}, \mathbf{a}) = -\mathbf{J}(\mathbf{a})^{\mathrm{T}} \cdot \mathbf{\Delta} \cdot \mathbf{J}(\mathbf{a}), \tag{13a}$$

$$\mathbf{J}(\mathbf{a}) \triangleq \frac{1}{\sqrt{2}} \left[\mathbf{I} - \mathbf{1}_N \cdot \mathbf{a}^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{13b}$$

where $\mathbf{1}_N$ is a vector whose entries are all one and $\mathbf{a}$ is a *signed distribution* of N terms such that $\sum_{i=1}^{N} a_i = 1$ and whose values influence the *origin* used to represent $\mathbf{X}$ [41].[23]

Recall Schoenberg's embedding theorem for Euclidean distance matrices [122] which states that the dissimilarity matrix $\mathbf{\Delta} \in \mathbb{D}_N(\eta)$ if and only if $\mathcal{K}(\mathbf{\Delta}, \mathbf{a}) \in \mathbb{S}_N(\eta)$, where $\mathbb{S}_N(\eta)$ is the space of symmetric PSD matrices with rank at most η.[24]

Then the metric scaling problem [37] can be seen as the minimization of the distance between the closed convex set of dissimilarity matrices and the set of symmetric positive matrices, which can be expressed as the following

[23] Notice that equation (13a) is simply the generalization of equation (12), *i.e.* the cosine-law, to all points of $\mathbf{X}$.

[24] Equation (13) is a linear mapping from $\mathbb{D}_N(\eta)$ to $\mathbb{S}_N(\eta)$.

optimization problem

$$\begin{aligned} \min_{\mathbf{G}} \quad & \|\mathcal{K}(\boldsymbol{\Delta}, \mathbf{a}) - \mathbf{G}\|_{\mathrm{F}}^2 , \\ \text{s.t.} \quad & \boldsymbol{\Delta} \in \mathbb{E}_N \\ & \mathbf{G} \in \mathbb{G}_N \end{aligned} \tag{14}$$

where $\mathbb{E}_N$ is the closed convex set of possible dissimilarity matrices and $\mathbb{G}_N$ a subset of $\mathbb{S}_N$ [47].

In the special case of a single dissimilarity matrix, namely $\mathbb{E}_N = \{\boldsymbol{\Delta}\}$, the optimization problem in equation (14) modifies into

$$\begin{aligned} \min_{\mathbf{G}} \quad & \|\mathcal{K}(\boldsymbol{\Delta}, \mathbf{a}) - \mathbf{G}\|_{\mathrm{F}}^2 , \\ \text{s.t.} \quad & \mathbf{G} \in \mathbb{S}_N(\eta), \end{aligned} \tag{15}$$

where the objective to be minimized is the STRAIN cost function in metric MDS [47].

As shown in [46] and further generalized in [47], an attractive feature of STRAIN is that its global solution can be computed explicitly. Moreover, accordingly to Schoenberg's embedding theorem $\mathcal{D}(\mathbf{X}) = \boldsymbol{\Delta}$ if and only if $\mathcal{K}(\boldsymbol{\Delta}, \mathbf{a})$ is a Gram matrix (*i.e*, $\mathbf{X} \cdot \mathbf{X}^{\mathrm{T}}$), from which it follows that $\mathbf{X}$ can be obtained by matrix factorization of a PSD matrix.

Indeed, let $\mathbf{K_a} = \mathcal{K}(\boldsymbol{\Delta}, \mathbf{a})$ be a kernel matrix constructed according to equation (13) for a particular value of $\mathbf{a}$, then the coordinate matrix $\mathbf{X}$ can be recovered – up to similarity transformation[25] – by

$$\mathbf{X} = [\mathbf{U_a}]_{1:N,1:\eta} \cdot [\boldsymbol{\Lambda_a}]_{1:\eta,1:\eta}^{\odot \frac{1}{2}} , \tag{16}$$

where $(\boldsymbol{\Lambda_a}, \mathbf{V_a})$ is the eigen-pair for $\mathbf{K_a}$ containing the decreasingly ordered eigenvalues and their corresponding eigenvectors respectively, and where $^{\odot m}$ denotes the m-th element-wise (Hadamard) power.

Some important generalizations of STRAIN are discussed in [47] while considerations about the robustness of STRAIN applied to metric MDS problems are offered in [123, 124]. More specifically it is shown that in the presence of dissimilarities corrupted by zero mean Gaussian noise STRAIN is comparable to a sub-optimal, in the ML-sense, cost function called SSTRESS.

[25] Similarity transformation includes scaling, rotation and shift and is recovered by Procrustes transformation [38].

In the particular case of $\mathbf{a} = \mathbf{1}_N/N$, then kernel matrix $\mathbf{G} \in \mathbb{S}_N(\eta)$ can be mapped into $\mathbf{D} \in \mathbb{D}_N(\eta)$ by the following linear operator

$$\mathcal{T}(\mathbf{G}) \triangleq \mathbf{1}_N \cdot \mathrm{diag}(\mathbf{G})^{\mathrm{T}} + \mathrm{diag}(\mathbf{G}) \cdot \mathbf{1}_N^{\mathrm{T}} - 2\mathbf{G}. \tag{17}$$

Also notice that that the inverse mapping is performed using the operator $\mathcal{K}(\boldsymbol{\Delta}, \mathbf{a})$ defined in equation (13) [50].

Classical multi-dimensional scaling

Consider the *double-centred* Euclidean kernel constructed from equation (13a) setting $\mathbf{a} = \frac{1}{N} \cdot \mathbf{1}_N$ and using all the M pairwise distances between the N points in $\mathbf{X}$ as [40]

$$\mathbf{K} = -\mathbf{J}^{\mathrm{T}} \cdot \mathbf{D} \cdot \mathbf{J}, \tag{18}$$

where

$$\mathbf{J} = \frac{1}{\sqrt{2}} \left[\mathbf{I}_N - \tfrac{1}{N} \cdot \mathbf{1}_N \mathbf{1}_N^{\mathrm{T}} \right]. \tag{19}$$

The C-MDS algorithm recovers the estimated configuration of points $\hat{\mathbf{X}}$ from $\mathbf{K}$ using equation (16). This solution represent the best approximation of $\mathbf{K}$ [46] that can be obtained under the constraint of $\mathrm{rank}(\mathbf{K}) = \eta$ and it is exact provided that $\mathbf{D} \in \mathbb{D}_N(\eta)$. In practical localization applications, however, one can only obtain a dissimilarity matrix $\tilde{\boldsymbol{\Delta}}$, corrupted by noise and bias[26], with the obvious consequence that no optimal close form solution is possible and where the value for the STRAIN cost function

$$\mathcal{C}_T = \| \hat{\mathbf{X}} \cdot \hat{\mathbf{X}}^{\mathrm{T}} - \mathcal{K}(\tilde{\boldsymbol{\Delta}}, \mathbf{a}) \|_{\mathrm{F}}^2$$

is a measure of the quality of the approximation obtained by the spectral truncation in equation (16).

Moreover using equation (17) it follows that

$$[\tilde{\mathbf{K}}]_{ij} = -\frac{1}{2} \left([\tilde{\mathbf{D}}]_{ij} - \mathbf{1}_N \cdot \sum_i \mathbf{a}_i [\tilde{\mathbf{D}}]_{ij} - \mathbf{1}_N \cdot \sum_j \mathbf{a}_j [\tilde{\mathbf{D}}]_{ij} + \sum_i \sum_j \mathbf{a}_i \mathbf{a}_j [\tilde{\mathbf{D}}]_{ij} \right). \tag{20}$$

From above is clear that the Euclidean kernel $\tilde{\mathbf{K}}$ used in the C-MDS solution is structured in such a way that each one of its elements depends on all the entries of $\tilde{\mathbf{D}}$ simultaneously, resulting in an error-propagation-prone algorithm.

[26] As discussed later, erasures may also occur.

Some interesting relations between the C-MDS algorithm and solutions based on principal component analysis (PCA) can be found in [36, 37].

3.1.2 The STRESS formulation

The other formulation commonly used in MDS, often referred to as the *closest Euclidean distance matrix problem* [50], which measures directly the fit between the dissimilarity matrix $\mathbf{\Delta}$ and an EDM,

$$\begin{aligned} \min_{\mathbf{D}} \quad & \left\| \mathbf{W}^{\odot 1/2} \circ (\mathbf{D}^{\odot r} - \mathbf{\Delta}^{\odot r}) \right\|_{\mathrm{F}}^2, \\ \text{s.t.} \quad & \mathbf{D} \in \mathbb{D}_N(\eta) \end{aligned} \tag{21}$$

where $\mathbf{W} \in \mathbb{R}^{N \times N}$ is a weight matrix whose entries account for the confidence on the corresponding dissimilarities in $\mathbf{\Delta}$ and $\circ$ denotes the Hadamard product.

As discussed in [125], the constraint in equation (21) is usually removed by exploiting equation (10), which has the effect of changing the objective function from

$$\min_{d} \quad \sum_{i<j} w_{ij} \left[(d_{ij}^2)^r - (\delta_{ij}^2)^r \right]^2, \tag{22}$$

into

$$\min_{\mathbf{x}} \quad \sum_{i<j} w_{ij} \left[\left(\sum_{k=1}^{\eta} (\mathbf{x}_{ik} - \mathbf{x}_{jk})^2 \right)^r - (\delta_{ij}^2)^r \right]^2. \tag{23}$$

Clearly, different cost functions result from different values of r in equation (23). Specifically, for $r = 0.5$ the objective function commonly referred to as the STRESS cost function in the MDS literature corresponds to the least squares on the distances. For specific values of w_{ij}, it was shown that STRESS is equivalent to the ML formulation of the localization problem [65]. In contrast, when $r = 1$ the objective function computes the least squares on the squared distances, and it is known as the SSTRESS cost function.

Both loss functions mentioned above were brought to the attention of the MDS community in the context of non-metric MDS by Kruskal in [126] and de Leeuw in [127].

While SSTRESS is generally everywhere smooth [125], it is often characterized by a large number of local minima and tends to favor large distances [38]. This,

together with the fact that in the localization problem affected by Gaussian noise SSTRESS is only a *sub-optimal* solution in the ML sense [65, 128, 129], makes STRESS cost function often preferable in the context of localization problems.

Concluding, it is important to notice that while the cost function in equation (14) solves the embedding problem by minimizing a least square objective function on inner product matrices, the formulation in equation (22) operates directly on the single distances d_{ij} and solves the closest-EDM problem.

The SMACOF algorithm

Although many standard minimization techniques could be used to solve STRESS, below we consider a solution based on the majorization algorithm initially proposed by Jan de Leeuw [130] and known under the name of SMACOF.

The idea is to minimize not the cost function $f_C(x)$, but a function $f_M(x, y)$ such that, except for the *support point* y in the domain of f_M where the two functions coincide, for any other generic point x in the domain of f_C, f_M is a majorizer of $f_C(x)$, namely

$$\begin{cases} f_C(x) \leqslant f_M(x, y), \\ f_C(y) = f_M(y, y). \end{cases} \tag{24}$$

If the conditions above are satisfied f_C can be solved through the majorizing function f_M. In context of metric MDS problems this majorization is obtained starting from the STRESS function $\mathcal{C}_S$ defined in equation (23) for $r = 0.5$, here rewritten as

$$\mathcal{C}_S(\mathbf{X}) = \eta_\delta^2 + \eta_d^2(\mathbf{X}) - 2\rho(\mathbf{X}), \tag{25}$$

where $\eta_\delta^2 = \sum_{i \leqslant j} w_{ij}\delta_{i,j}^2$ and $\eta_d^2(\mathbf{X}) = \sum_{i \leqslant j} w_{ij} d_{ij}^2(\mathbf{X})$ while the last term is $\rho(\mathbf{X}) = \sum_{i \leqslant j} w_{ij}\delta_{ij} d_{ij}(\mathbf{X})$.

Rewriting equation (25) in matrix form yields [38]

$$\mathcal{C}_S(\mathbf{X}) = \eta_\delta^2 + \operatorname{trace}\left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{H_W} \cdot \mathbf{X}\right) - 2 \cdot \operatorname{trace}\left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{P_S}(\mathbf{X}) \cdot \mathbf{X}\right), \tag{26}$$

where $\mathbf{H_W}$ is such that

$$h_{\mathbf{W},ij} = \begin{cases} -w_{ij} & i \neq j, \\ \sum_{k \neq i} w_{ik} & i = j, \end{cases} \tag{27}$$

and $\mathbf{P_S}$ is such that

$$p_{\mathbf{S},ij} = \begin{cases} -w_{ij}\delta_{ij}d_{ij}^{-1}\left(\mathbf{X}\right) & i \neq j \qquad d_{ij}\left(\mathbf{X}\right) \neq 0, \\ 0 & i \neq j \qquad d_{ij}\left(\mathbf{X}\right) = 0, \\ -\sum_{k\neq i} p_{\mathbf{S},ik} & i = j. \end{cases} . \tag{28}$$

As shown in [38, 131] a majorized form of equation (26) is

$$\mathcal{C}_M\left(\mathbf{X},\mathbf{Y}\right) = \eta_\delta^2 + \text{trace}\left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{H_W} \cdot \mathbf{X}\right) - 2 \cdot \text{trace}\left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{P_S}\left(\mathbf{Y}\right) \cdot \mathbf{Y}\right). \tag{29}$$

The function above is minimized by iteratively equating its gradient, computed with respect to the matrix $\mathbf{Y}$, to 0, yielding

$$\frac{\partial \mathcal{C}_M}{\partial \mathbf{Y}} = 2\mathbf{H_W} \cdot \mathbf{X} - 2\mathbf{P_S}\left(\mathbf{Y}\right) \cdot \mathbf{Y} = 0. \tag{30}$$

It follows that the solution at the iteration $[\bar{i}+1]$ is obtained from the previous one from what is commonly known as the Guttman's transform, namely

$$\mathbf{X}_{[\bar{i}+1]} = \mathbf{H}_{\mathbf{W}}^{\dagger} \cdot \mathbf{P_S}\left(\mathbf{Y}_{[\bar{i}]}\right) \cdot \mathbf{Y}_{[\bar{i}]}, \tag{31}$$

where † indicates the Moore-Penrose pseudo-inverse.

Although different and more complex optimizations, such as *Newtonian* or *Quasi-Newtonian* solutions [125, 131] could be used to solve STRESS, the majorization employed in SMACOF has the advantage of generating convex approximations of the original function, which in turn implies non-decreasing sequences of the values for the cost function, making the solution more robust to local minima [38]. Motivated by this reason, the SMACOF algorithm has been used to solve multitarget localization problems with centralized and distributed system architectures [69].

In addition to the above, in [132] it is shown that in many cases the search direction for the SMACOF algorithm can be inferred from the previous iterations, giving the possibility to add a vector extrapolation stage inside the algorithm to accelerate the convergence time of overall optimization.[27]

[27] Bear in mind that vector extrapolation techniques do not ensure convexity and for this reason they usually incorporate a check on values of the cost function to ensure the convergence of the optimization to a minima.

3.1.3 The Bayesian formulation

Differently from above where the STRAIN and the STRESS formulation were introduced to cope with parameter estimation problems commonly encountered in localization applications, in the following STT scenarios are modeled as *discrete-time filtering* problems which can be optimally solved, in a Bayesian sense, through the *filtering distribution*.

Let $\bar{k} \in \mathbb{N}$ be the discrete time index assigned to the sampling interval $\bar{T}_{\bar{k}-1} \triangleq t_{\bar{k}} - t_{\bar{k}-1}$ and let also the target state vector $\boldsymbol{x}_{\bar{k}} \in \mathbb{R}^{n_x}$ with initial probability $p(\boldsymbol{x}_0)$ be a *Markov chain*. Under the assumption $\boldsymbol{x}_{\bar{k}}$ depends only on $\boldsymbol{x}_{\bar{k}-1}$, which translates into the two following properties:

1. *Independence from the past*:

$$p(\boldsymbol{x}_{\bar{k}}|\boldsymbol{x}_{1:\bar{k}-1}, \boldsymbol{z}_{1:\bar{k}-1}) = p(\boldsymbol{x}_{\bar{k}}|\boldsymbol{x}_{\bar{k}-1}),$$

2. *Independence of the past from the future given the current state*:

$$p(\boldsymbol{x}_{\bar{k}-1}|\boldsymbol{x}_{\bar{k}:\bar{k}+\bar{T}}, \boldsymbol{z}_{\bar{k}:\bar{k}+\bar{T}}) = p(\boldsymbol{x}_{\bar{k}-1}|\boldsymbol{x}_{\bar{k}}).$$

In addition to the Markovian property of the state vector, defied by the two properties above, it is common to assume conditional independence between the observations $\boldsymbol{z}_{1:\bar{k}}$ and between observations and the state history $\boldsymbol{x}_{1:\bar{k}}$, yielding

$$p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}_{1:\bar{k}}, \boldsymbol{z}_{1:\bar{k}-1}) = p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}_{\bar{k}}). \tag{32}$$

Under the assumptions above, the system is said to evolve as the hidden Markov model (HMM) illustrated in Figure 8, and it can be represented by the discrete-time[28] system model:

$$\boldsymbol{x}_{\bar{k}|\bar{k}-1} \sim p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{x}_{\bar{k}-1}), \tag{33}$$

$$\boldsymbol{z}_{\bar{k}} \sim p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}_{\bar{k}}), \tag{34}$$

where $p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{x}_{\bar{k}-1})$ defines the *prior* knowledge on the model and the *likelihood* function $p_{\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}_{\bar{k}})$ represents the knowledge on the measurement process.[29]

[28] Although not considered in this work, models for *continuous-discrete* and *continuous-time* systems are known in the literature and can be found in [30, 133] while recent advances in the area are offered in [134].

[29] Implicitly it also defines the noise statistics.

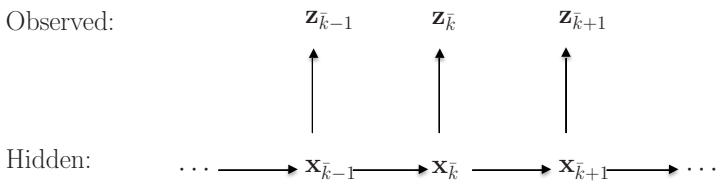


Fig 8. Representation of the HMM commonly employed in recursive systems.

As a consequence of the assumption on the system model, the joint prior distributions and the joint likelihood for the system are computed as

$$p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{0:\bar{k}}) = \quad p(\boldsymbol{x}_0)\prod_{i=1}^{\bar{k}} p_{\bar{i}|\bar{i}-1}(\boldsymbol{x}_{\bar{i}}|\boldsymbol{x}_{\bar{i}-1}), \tag{35a}$$

$$p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{1:\bar{k}}|\boldsymbol{x}_{0:\bar{k}}) = \quad \prod_{i=1}^{k} p_{\bar{i}}(\boldsymbol{z}_{\bar{i}}|\boldsymbol{x}_{\bar{i}}). \tag{35b}$$

If follows that the joint posterior density $p_{\bar{k}}(\boldsymbol{x}_{1:\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$ is obtained applying the Bayes' rule as

$$p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{1:\bar{k}}|\boldsymbol{z}_{1:\bar{k}}) = \frac{p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{1:\bar{k}}|\boldsymbol{x}_{1:\bar{k}})p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{1:\bar{k}})}{p_{\bar{k}|\bar{k}-1}(\boldsymbol{z}_{1:\bar{k}})} \propto p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{1:\bar{k}}|\boldsymbol{x}_{1:\bar{k}})p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{1:\bar{k}}). \tag{36}$$

Although $p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{1:\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$ represents the conceptual solution to problem at time $\bar{k}$, it suffers from an increase in computation time. Therefore it is common to replace the full posterior distribution by the *filtering* distribution[30] $p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$. This is particularly useful in LT applications where it is of interest to recursively quantify the belief in $\boldsymbol{x}_{\bar{k}}$ given $\boldsymbol{z}_{1:\bar{k}}$ and which is obtained through the *marginal posterior* distribution $p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$ [30]. Thus, assuming[31] $p_{\bar{k}-1}(\boldsymbol{x}_{\bar{k}-1}|\boldsymbol{z}_{1:\bar{k}-1})$ known, then the filtering distribution at $\bar{k}$ can be computed using the discrete-time Bayesian, that is:

– *Prediction*: use the *Chapman-Kolmogorov* equation and the system model

[30]The following marginals can also be computed:

– *prediction* distribution $p_{\bar{k}+\bar{n}|\bar{k}}(\boldsymbol{x}_{\bar{k}+\bar{n}}|\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{\bar{k}})$, marginal of the future states;
– *smoothing* distribution $p_{\bar{i}|1:\bar{k}}(\boldsymbol{x}_{\bar{i}}|\boldsymbol{z}_{1:\bar{k}})$, marginal of $\boldsymbol{x}_{\bar{i}}$ given the set $\boldsymbol{z}_{1:\bar{k}}$ where $1 \leqslant i \leqslant k$.

[31]The distribution $p_0(\boldsymbol{x}_0) \triangleq p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1})$ at $\bar{k}=0$ is assumed known.

defined in equation (35) to compute the prediction density as

$$p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1}) = \int_{\mathbb{R}^{n_{\mathbf{x}}}} p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{x}) p_{\bar{k}-1}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}-1}) d\boldsymbol{x}. \tag{37}$$

– *Update*: use the Bayes' rule to correct on the basis of the current observation and the statistical characterization of $\boldsymbol{z}_{\bar{k}}$:

$$\begin{aligned} p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}}) &= p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}}, \boldsymbol{z}_{1:\bar{k}-1}) \\ &= \frac{p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}_{\bar{k}}, \boldsymbol{z}_{1:\bar{k}-1}) p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1})}{p_{\bar{k}|\bar{k}-1}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1})} \\ &= \frac{p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}_{\bar{k}}) p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1})}{p_{\bar{k}|\bar{k}-1}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1})}, \end{aligned} \tag{38}$$

where the normalizing constant is

$$p_{\bar{k}|\bar{k}-1}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1}) = \int_{\mathbb{R}^{n_{\mathbf{x}}}} p_{\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}) p_{\bar{k}-1}(\boldsymbol{x}|\boldsymbol{z}_{1:\bar{k}-1}) d\boldsymbol{x}. \tag{39}$$

Once $p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$ is known, the optimal state estimate is obtained using the predefined decision criterion, for instance the MMSE or the MAP estimator mentioned in Section 2.3.2. Similarly, a measure of accuracy of a state estimate (*e.g.*, covariance) may also be obtained from $p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$.

Alternatively, when target motion and measurement statistic are characterized by a transition density $p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{x}_{\bar{k}-1})$ and a likelihood function $p_{\bar{k}|\bar{k}}(\boldsymbol{z}_{\bar{k}}|\boldsymbol{x}_{\bar{k}})$, the system can be described using the following *discrete-time stochastic* model

$$\boldsymbol{x}_{\bar{k}} = f\left(\boldsymbol{x}_{\bar{k}-1}, \boldsymbol{q}_{\bar{k}-1}\right), \qquad \text{(Dynamic Model)} \tag{40a}$$

$$\boldsymbol{z}_{\bar{k}} = g\left(\boldsymbol{x}_{\bar{k}}, \boldsymbol{r}_{\bar{k}}\right), \qquad \text{(Measurement Model)} \tag{40b}$$

where $f(\cdot)$ and $g(\cdot)$ are known, possibly nonlinear functions, used to describe the motion and the measurement model of $\boldsymbol{x}_{\bar{k}-1}$ while $\boldsymbol{q}_{\bar{k}-1}$ and $\boldsymbol{r}_{\bar{k}}$ are the process and measurement noise sequence used to account for miss-modeling effects and disturbances in systems. Let $\boldsymbol{x}_{\bar{k}|\bar{k}-1}$ A representation of the recursive estimate of the filtering distribution for the system model in equation (40) is provide in Figure 9.

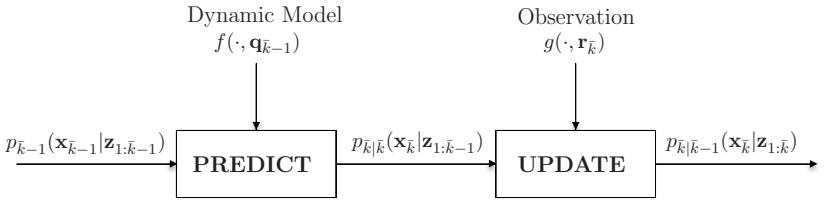


Fig 9. Representation of the Bayesian recursion used to compute the filtering distribution.

However, the optimal filtering recursion described above is theoretical since only in a limited number of cases it can be computed analytically.[32]

Such cases are [77]:

- LG models
- discrete state space with a finite number of states
- few other subclasses of nonlinear problems different from the ones above [30].

In all other cases, which include the most practical situations, it is necessary to rely either on approximations or suboptimal Bayesian algorithms.

Kalman filters

Let $f(\cdot)$ and $g(\cdot)$ in equation (40) be two known linear functions of $\{\boldsymbol{x}_{\bar{k}-1}, \boldsymbol{q}_{\bar{k}-1}\}$ and $\{\boldsymbol{z}_{\bar{k}}, \boldsymbol{r}_{\bar{k}}\}$ respectively with $\boldsymbol{q}_{\bar{k}-1}$ and $\boldsymbol{r}_{\bar{k}}$ as mutually independent zero-mean white Gaussian distributions whose covariance matrices are $\boldsymbol{Q}_{k-1}$ and $\boldsymbol{R}_k$ respectively. Then the system in equation (40) can be written as

$$\boldsymbol{x}_{\bar{k}} = \mathbf{F}_{\bar{k}-1} \cdot \boldsymbol{x}_{\bar{k}-1} + \boldsymbol{q}_{\bar{k}-1}, \tag{41a}$$

$$\boldsymbol{z}_{\bar{k}} = \boldsymbol{G}_{\bar{k}} \cdot \boldsymbol{x}_{\bar{k}} + \boldsymbol{r}_{\bar{k}}, \tag{41b}$$

where $\mathbf{F}_{\bar{k}-1} \in \mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ and $\boldsymbol{G}_{\bar{k}} \in \mathbb{R}^{n_{\boldsymbol{z}} \times n_{\boldsymbol{x}}}$ are known matrices.

Under the assumption that $p_0(\boldsymbol{x}_{\bar{0}})$ is Gaussian, it follows that $p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$ is also Gaussian and therefore completely characterized by its first two moments, mean and covariance [30]. Using equations (37) and (38), the relations for the

[32] Its computation would require the entire filtering pdf obtained solving the multidimensional integrals in equation (37) and (39).

filtering distribution are

$$p_{\bar{k}-1}(\boldsymbol{x}_{\bar{k}-1}|\boldsymbol{z}_{1:\bar{k}-1}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{x}_{\bar{k}-1}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}-1}), \tag{42}$$

$$p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}-1}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{x}_{\bar{k}|\bar{k}-1}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}), \tag{43}$$

$$p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{x}_{\bar{k}|\bar{k}}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}}), \tag{44}$$

where $\mathcal{N}(\boldsymbol{x}; \mathbf{m}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{x}})$ is a Gaussian distribution with mean $\mathbf{m}_{\boldsymbol{x}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ is the Cholesky factorization of the state covariance matrix, *i.e.*, $\mathbf{P}_{\boldsymbol{x}} = \boldsymbol{\Sigma}_{\boldsymbol{x}} \cdot (\boldsymbol{\Sigma}_{\boldsymbol{x}})^{\mathrm{T}}$.

In particular, the parameters characterizing the filtering distributions above are computed applying the Bayesian recursion as:
Step 1 - Prediction

$$\boldsymbol{m}_{\bar{k}|\bar{k}-1} = \mathbf{F}_{\bar{k}-1} \cdot \boldsymbol{x}_{\bar{k}-1}, \tag{45}$$

$$\mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1} = \boldsymbol{Q}_{\bar{k}-1} + \mathbf{F}_{\bar{k}-1} \cdot \mathbf{P}_{\boldsymbol{x},\bar{k}-1} \cdot \mathbf{F}_{\bar{k}-1}^{\mathrm{T}}. \tag{46}$$

Step 2 - Update

$$\boldsymbol{m}_{\bar{k}|\bar{k}} = \boldsymbol{m}_{\bar{k}|\bar{k}-1} + \boldsymbol{K}_{\bar{k}} \cdot (\boldsymbol{z}_{\bar{k}} - \boldsymbol{G}_{\bar{k}} \cdot \boldsymbol{m}_{\bar{k}|\bar{k}-1}), \tag{47}$$

$$\mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}} = \mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1} - \boldsymbol{K}_{\bar{k}} \cdot \mathbf{S}_{\bar{k}|\bar{k}-1} \cdot \boldsymbol{K}_{\bar{k}}^{\mathrm{T}}, \tag{48}$$

where

$$\mathbf{S}_{\bar{k}} = \boldsymbol{G}_{\bar{k}} \cdot \mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1} \cdot \boldsymbol{G}_{\bar{k}}^{\mathrm{T}} + \boldsymbol{R}_{\bar{k}}, \tag{49}$$

is the covariance of the innovation term $\boldsymbol{y}_{\bar{k}} = \boldsymbol{z}_{\bar{k}} - \boldsymbol{G}_{\bar{k}} \cdot \boldsymbol{x}_{\bar{k}|\bar{k}-1}$, and

$$\boldsymbol{K}_{\bar{k}} = \mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1} \cdot \boldsymbol{G}_{\bar{k}}^{\mathrm{T}} \cdot \mathbf{S}_{\bar{k}}^{-1}, \tag{50}$$

is the *gain* matrix.[33]

At the following iteration the state vector $\boldsymbol{x}$ is then set equal to the previous estimate of $\boldsymbol{m}$.

The recursion presented above is the well known KF [78, 87] which is proved to be optimal in the presence of LG problems.[34] The KF can also be used as a suboptimal solution to the more general case of nonlinear problems in the form of EKF [30], UKF [83] and CKF [84] based filters.

[33] Also notice that $\mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}} = [\mathbf{I} - \boldsymbol{K}_{\bar{k}} \cdot \boldsymbol{G}_{\bar{k}}] \cdot \mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1}$.
[34] The derivation of the KF equations can be obtained using a LS argument [30].

Gaussian mixture filters

The main limitation of the KF approach mentioned above is the Gaussian assumption implicit in the filtering distribution $p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}}|\boldsymbol{z}_{1:\bar{k}})$. To overcome this problem and extend the approach to more general scenarios, *e.g.* pdfs characterized by multimodal distributions, it was proposed to represent a generic filtering distributions by the sum of Gaussian distributions [79].

This is in agreement with Wiener's Theorem on approximation which proves that any distribution can be approximated up to an arbitrary degree by a sum of Gaussian distribution [120]. Let the filtering distribution at $\bar{k}-1$ is given in the form of the following GM

$$p_{\bar{k}-1}(\boldsymbol{x}_{\bar{k}-1}|\boldsymbol{Z}^{\bar{k}-1}) = \sum_{j=1}^{J_{\bar{k}-1}} \omega_{\bar{k}-1}^{(j)} \cdot \mathcal{N}(\boldsymbol{x}; \mathbf{m}_{\bar{k}-1}^{(j)}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)}), \tag{51}$$

such that $\sum_{j=1}^{J_{\bar{k}}-1} \omega_{\bar{k}-1}^{(j)} = 1$.

Then, for LG system models the filtering distribution at $\bar{k}$ can be computed using the Bayesian paradigm, namely:

Step 1 - Prediction

$$p_{\bar{k}|\bar{k}-1}(\boldsymbol{x}_{\bar{k}|\bar{k}-1}|\boldsymbol{z}_{1:\bar{k}-1}) = \sum_{j=1}^{J_{\bar{k}-1}} \omega_{\bar{k}-1}^{(j)} \cdot \mathcal{N}(\boldsymbol{x}; \mathbf{m}_{\bar{k}|\bar{k}-1}^{(j)}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}^{(j)}), \tag{52}$$

where $\mathbf{m}_{\bar{k}|\bar{k}-1}^{(j)}$ and $\boldsymbol{\Sigma}_{\bar{k}|\bar{k}-1}^{(j)}$ are computed using the KF prediction equation (45) and (46).

Step 2 - Update

$$p_{\bar{k}|\bar{k}}(\boldsymbol{x}_{\bar{k}|\bar{k}}|\boldsymbol{z}_{1:\bar{k}}) = \sum_{j=1}^{J_{\bar{k}}} \omega_{\bar{k}}^{(j)} \cdot \mathcal{N}(\boldsymbol{x}; \mathbf{m}_{\bar{k}|\bar{k}}^{(j)}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}}^{(j)}), \tag{53}$$

where $\mathbf{m}_{\bar{k}|\bar{k}}^{(j)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}}^{(j)}$ are computed using the KF update equation (47) and (48) and

$$\omega_{\bar{k}}^{(j)} = \omega_{\bar{k}-1}^{(j)} \cdot \frac{\mathcal{N}(\boldsymbol{x}; \boldsymbol{G}_{\bar{k}} \cdot \mathbf{m}_{\bar{k}|\bar{k}}^{(j)}, \boldsymbol{G}_{\bar{k}} \cdot \mathbf{P}_{\bar{k}|\bar{k}-1}^{(j)} \cdot \boldsymbol{G}_{\bar{k}}^{\mathrm{T}} + \boldsymbol{R}_{\bar{k}})}{\sum_{i=1}^{J_{\bar{k}}} \mathcal{N}(\boldsymbol{x}; \boldsymbol{G}_{\bar{k}} \cdot \mathbf{m}_{\bar{k}|\bar{k}}^{(i)}, \boldsymbol{G}_{\bar{k}} \cdot \mathbf{P}_{\bar{k}|\bar{k}-1}^{(i)} \cdot \boldsymbol{G}_{\bar{k}}^{\mathrm{T}} + \boldsymbol{R}_{\bar{k}})}. \tag{54}$$

Once the filtering distribution at $\bar{k}$ is available, the state variable and the associated covariance matrix are found as:

$$\boldsymbol{x}_{\bar{k}|\bar{k}} = \sum_{j=1}^{J_{\bar{k}}} \omega_{\bar{k}}^{(j)} \mathbf{m}_{\bar{k}|\bar{k}}^{(j)}, \tag{55}$$

$$\mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}} = \sum_{j=1}^{J_{\bar{k}}} \omega_{\bar{k}}^{(j)} \left[\mathbf{P}_{\boldsymbol{x},\bar{k}}^{(j)} + (\boldsymbol{x}_{\bar{k}|\bar{k}} - \mathbf{m}_{\bar{k}|\bar{k}}^{(j)}) \cdot (\boldsymbol{x}_{\bar{k}|\bar{k}} - \mathbf{m}_{\bar{k}|\bar{k}}^{(j)})^{\mathrm{T}} \right]. \tag{56}$$

3.2 The MTT problem with passive sensors

In the STT scenario the filter is constructed under the knowledge that a single target exists, such that only the state vector and measurements vary in time. In contrast, in MTT scenarios, also the number of targets change with time. Furthermore, typically there is no model to govern the likelihood with which the targets appear and fade from the surveillance region.

In order to cope with this difficulty specific to the MTT scenario, the standard Bayesian framework associated with the conventional Kalman filter briefly discussed above needs to be replaced by more flexible frameworks. Such a framework was suggested by Mahler, who used the FISST tools to extend the Bayesian formulation to the MTT scenarios [103].

3.2.1 The Bayesian formulation with random finite sets

Among other possibilities, the recently emerged Random Finite Set (RFS) framework [116] is of particular interest since it requires no assumption on the number of targets, unlike more established alternatives such as the JPDA. Differently from the MHT approach, does not rely on the idea of hypothesizing associations between measurements and targets [121, pp. 8].

Let $\mathcal{X}$ and $\mathcal{Z}$ denote the *state* and the *observation* spaces, respectively. Let also the RFS $\boldsymbol{X}_{\bar{k}}$ describe the set of state vectors corresponding to $L_{\bar{k}}$ targets, and the RFS $\boldsymbol{Z}_{\bar{k}}$ describe the $M_{\bar{k}}$ observations at instant $\bar{k}$, *i.e.*

$$\boldsymbol{X}_{\bar{k}} = \{\boldsymbol{x}_{\bar{k},1}, \ldots, \boldsymbol{x}_{\bar{k},L_k}\} \in F(\mathcal{X}), \tag{57}$$

$$\boldsymbol{Z}_{\bar{k}} = \{\boldsymbol{z}_{\bar{k},1}, \ldots, \boldsymbol{z}_{\bar{k},M_k}\} \in F(\mathcal{Z}), \tag{58}$$

where $F(\mathcal{X})$ is the space of all finite subsets of $\mathcal{X}$ and $F(\mathcal{Z})$ is the space of all

finite subsets of $\mathcal{Z}$.

Both $\boldsymbol{X}_{\bar{k}}$ and $\boldsymbol{Z}_{\bar{k}}$ are described by a discrete distribution and a joint probability density function characterizing the number of elements of the sets as well as the distribution of their states. Following the model suggested in [118], the RFS for the target states is given by

$$\boldsymbol{X}_{\bar{k}} = \left[\bigcup_{\boldsymbol{\xi} \in \boldsymbol{X}_{\bar{k}|\bar{k}-1}} \boldsymbol{E}_{\bar{k}|\bar{k}-1}(\boldsymbol{\xi}) \right] \cup \left[\bigcup_{\boldsymbol{\xi} \in \boldsymbol{X}_{\bar{k}|\bar{k}-1}} \boldsymbol{B}_{\bar{k}|\bar{k}-1}(\boldsymbol{\xi}) \right] \cup \boldsymbol{\Gamma}_{\bar{k}}, \tag{59}$$

where $\boldsymbol{E}_{\bar{k}|\bar{k}-1}(\boldsymbol{\xi})$ is the RFS describing the evolved (or the prediction of surviving) targets $\boldsymbol{x}_{\bar{k}|\bar{k}-1}$ in time, $\boldsymbol{B}_{\bar{k}|\bar{k}-1}(\boldsymbol{\xi})$ is the RFS describing the targets that spawn from previous trajectories $\boldsymbol{\xi}_{\bar{k}|\bar{k}-1}$ and $\boldsymbol{\Gamma}_{\bar{k}}$ an additional RFS included to account for new target births at instant $\bar{k}$.

Similarly the RFS describing the observations can be expressed by

$$\boldsymbol{Z}_{\bar{k}} = \left[\bigcup_{\boldsymbol{x} \in \boldsymbol{X}_{\bar{k}}} \boldsymbol{M}_{\bar{k}}(\boldsymbol{x}) \right] \cup \boldsymbol{H}_{\bar{k}}, \tag{60}$$

where $\boldsymbol{H}_{\bar{k}}$ accounts for the clutter measurements and $\boldsymbol{M}_{\bar{k}}$ models the measurements generated by $\boldsymbol{x}_{\bar{k}}$ conditioned on the detection probability p_{D}.

Using the definition of set integral and set derivative from FISST theory the probability density function for the RFS is

$$p_{\bar{k}}\left(\{\boldsymbol{x}_{\bar{k}}^{(1)}, \ldots, \boldsymbol{x}_{\bar{k}}^{(n)}\}|n\right) = n! p_{\bar{k}}\left(\boldsymbol{x}_{\bar{k}}^{(1)}, \ldots, \boldsymbol{x}_{\bar{k}}^{(n)}\right), \tag{61}$$

where the factorial term n is a direct consequence of the definition of set integral in FISST and indicates the probability of the existence of the elements of $\boldsymbol{X}_{\bar{k}}$ independently on their possible permutations. As already mentioned above, equation (61) shows that the RFS $\boldsymbol{X}_{\bar{k}}$ is described by $p_{\bar{k}}\left(\boldsymbol{X}_{\bar{k}}\right)$ and a cardinality distribution from now on denoted by $p_{\mathrm{C},\bar{k}|\bar{k}}(\cdot)$. Therefore, bearing equations (59) and (60) in mind and relying on the FISST tool an analogous to the Bayesian recursion for the STT problem the conceptual solution to the MTT case becomes [113, 116, pp. 14]

$$p_{\bar{k}|\bar{k}-1}(\boldsymbol{X}_{\bar{k}}|\boldsymbol{Z}_{1:\bar{k}}) = \int \varphi_{\bar{k}|\bar{k}-1}(\boldsymbol{X}_{\bar{k}}|\boldsymbol{X})\, p_{\bar{k}-1}\big(\boldsymbol{X}|\boldsymbol{Z}_{1:\bar{k}-1}\big)\mu(d\boldsymbol{X})\,, \tag{62}$$

$$p_{\bar{k}}\left(\boldsymbol{X}_{\bar{k}}|\boldsymbol{Z}_{1:\bar{k}}\right) = \frac{h_{\bar{k}}\left(\boldsymbol{Z}_{\bar{k}}|\boldsymbol{X}_{\bar{k}}\right) p_{\bar{k}|\bar{k}-1}\left(\boldsymbol{X}_{\bar{k}}|\boldsymbol{Z}_{1:\bar{k}-1}\right)}{\int h_{\bar{k}}\left(\boldsymbol{Z}_{\bar{k}}|\boldsymbol{X}\right) p_{\bar{k}|\bar{k}-1}(\boldsymbol{X}_{\bar{k}}|\boldsymbol{Z}_{1:\bar{k}})\mu\left(d\boldsymbol{X}\right)}. \tag{63}$$

where $\varphi_{\bar{k}|\bar{k}-1}(\cdot|\cdot)$ and $h_{\bar{k}|\bar{k}-1}(\cdot|\cdot)$ are the multitarget transition density and likelihood function respectively.

A SMC recursion that propagates in time the entire MTT posterior pdf described in equation (63) was proposed in [117]. It was also discussed thereby, however, that such a solution becomes prohibitive in systems with more than a few targets, among other reasons, because equations (62) and (63) contain integrals with multiple and *variable* dimensions. A representation of this conceptual solution is provided in Figure 10. Bear in mind that since both Figure 9 and 10 are Bayesian solutions to the STT and MTT problem respectively, they both share the prediction and update stage. However, as shown in Figure 10, in presence of MTT problems the entire RFS used to represent $\boldsymbol{X}$ must be propagated forward in time by the recursion. Indeed, while in the STT case the posterior pdf of the single state vector is propagated in time, in the MTT case this is substituted by the posterior pdf of a RFS describing the evolution of the targets in time.

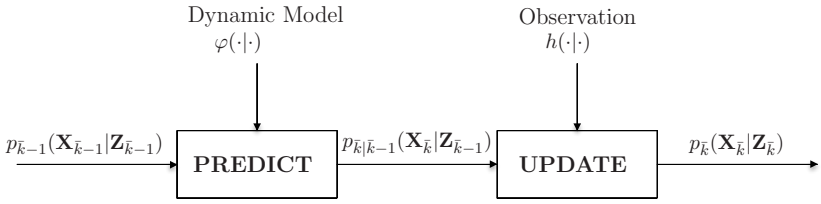


Fig 10. Representation of the Bayesian recursion used to compute the MTT filtering distribution.

The probability hypothesis density filter

The propagation in time of the entire MTT posterior pdf defined in equation (63) would be too computationally demanding for any practical system. Fortunately, a suboptimal solution was proposed in [116] where only the first order moment ν of the MTT posterior pdf – referred to as the PHD – is propagated. Specifically, under the assumptions that [118]:

– targets evolve and generate observations independently of one another
– clutter and target measurements are independent of each other
– the predicted multitarget RFS governed by $p_{k|k-1}$ is Poisson,

then the posterior intensity can be propagated in time by PHD recursion:

$$\begin{aligned} \nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}) &= \int p_{\mathrm{S},\bar{k}}(\boldsymbol{\xi})\, \varphi_{\bar{k}|\bar{k}-1}(\boldsymbol{x}|\boldsymbol{\xi})\, \nu_{\bar{k}-1}(\boldsymbol{\xi})\, d\boldsymbol{\xi} \\ &+ \int \beta_{\bar{k}|\bar{k}-1}(\boldsymbol{x}|\boldsymbol{\xi})\, \nu_{\bar{k}-1}(\boldsymbol{\xi})\, d\boldsymbol{\xi} + \gamma(\boldsymbol{\xi}), \end{aligned} \tag{64}$$

and

$$\begin{aligned} \nu_{\bar{k}|\bar{k}}(\boldsymbol{x}) &= \left[1 - p_{\mathrm{D},\bar{k}}(\boldsymbol{x})\right] \nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}) \\ &+ \sum_{\boldsymbol{z}\in\boldsymbol{Z}_{\bar{k}}} \frac{p_{\mathrm{D},\bar{k}}(\boldsymbol{x})\, h_{\bar{k}}(\boldsymbol{z}|\boldsymbol{x})\, \nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x})}{\kappa_{\bar{k}}(\boldsymbol{x}) + \int p_{\mathrm{D},\bar{k}}(\boldsymbol{\xi})\, h_{\bar{k}}(\boldsymbol{z}|\boldsymbol{\xi})\, \nu_{\bar{k}|\bar{k}-1}(\boldsymbol{\xi})\, d\boldsymbol{\xi}}, \end{aligned} \tag{65}$$

where $p_{\mathrm{S},\bar{k}}(\boldsymbol{\xi})$ is the probability of survival of the target in $\boldsymbol{\xi}$ at $\bar{k}$, $\beta_{\bar{k}|\bar{k}-1}(\cdot|\boldsymbol{\xi})$ and $\gamma(\cdot)$ respectively the intensities of the RFS $\boldsymbol{B}_{\bar{k}}$ and $\boldsymbol{\Gamma}_{\bar{k}}$ at $\bar{k}$, $p_{\mathrm{D},\bar{k}}(\boldsymbol{x})$ is the detection probability of $\boldsymbol{x}$ at $\bar{k}$ and $\kappa_{\bar{k}}(\cdot)$ is the intensity of the clutter RFS.

The PHD[35] is based on the first order approximation of the function that maps sets of targets into a probability space, namely Dirac delta at the targets' locations and it is characterized by the property that the cardinality of $\boldsymbol{X}_{\bar{k}}$ can then be estimated by

$$n_{\bar{k}|\bar{k}}(\boldsymbol{X}) = \int_{\mathcal{X}} \nu_{\bar{k}|\bar{k}}(\boldsymbol{x}) \cdot d\boldsymbol{x}. \tag{66}$$

One advantage of the PHD framework is that the MTT posterior pdf can (in principle) be evaluated via the Bayesian rule (equation (62) and (63)), estimates of $\boldsymbol{X}$ can be computed, similarly to the Bayesian STT case, under different *optimality* criteria.[36]

A further advantage is that the PHD intensity function $\nu_{\bar{k}|\bar{k}}$ is a function of the state vectors of individual targets[37], and therefore has a fixed number of dimensions, specifically as many as the size of the state vector $\boldsymbol{x}$.

Nevertheless, propagating $\nu_{\bar{k}|\bar{k}}$ may still be computationally challenging, since it is still a multivariate function which must be approximated accurately enough over the entire surveillance region. Initial attempts to mitigate this problem

[35] Since the PHD does not integrate to 1, it is not a density function.

[36] Depending on the algorithm used to implement the PHD recursion, different techniques can be employed to extract the states of the $n_{\bar{k}|\bar{k}}(\boldsymbol{X})$ targets. Specifically SMC implementations require a clustering stage [117], while GM-based approaches rely on the fact that the means of the GM components are the maxima of the intensity function [118].

[37] For details we refer readers to [116].

were based on SMC (particle) implementations [117], which may suffice in some cases but quickly becomes unfeasible as the number of targets increases.

To alleviate the computational complexity of SMC implementations of the PHD, a close-form solution of the recursion based on the GM of the MTT intensity function was proposed in [118]. This solution was shown to be optimum under the assumption that:

- target trajectories and sensors measurements follow linearGaussian models
- clutter measurements follow Poisson processes
- the probabilities of survival and detection are constant (*i.e.*, $p_{\mathrm{S},\bar{k}}(\xi) = p_{\mathrm{S},\bar{k}}$ and $p_{\mathrm{D},\bar{k}}(\boldsymbol{x}) = p_{\mathrm{D},\bar{k}}$)
- the intensity of the birth RFS is a GM.

Outside these assumptions, however, the GM implementation of the PHD filter is suboptimal as it relies on approximations such as linearization (in similarity to the EKF) or the unscented transform (in similarity to the UKF).

The cardinalized probability hypothesis density filter

One problem of the PHD recursion mentioned above is that, although the RFS $\boldsymbol{X}_{\bar{k}}$ is described by both the distribution of its elements as well as a cardinality distribution, in the PHD recursion the latter is not considered since the the number of targets at $\bar{k}$ is computed as the mean of the PHD function through equation (66). As explained in [119], this results in inconsistent estimates of the number of targets. To overcome this problem in [135] Mahler proposed to propagate in time the cardinality distribution $p_{\bar{k}}(n)$ for the number of targets together with the intensity function.

The assumptions at the core of the CPHD recursion are:

- targets evolve and generate observations independently of one another
- new-born and surviving targets are independent of each another[38]
- clutter measurements are an *i.i.d* cluster process
- prior and predicted multitarget RFS are i.i.d cluster processes.

Let $C_j^{\ell} = \ell!/j!(\ell - j)!$ be the binomial coefficient, and $\langle \cdot, \cdot \rangle$ the inner product between the two real valued functions given as argument, then the posterior

[38]The formulation can be extended to include target spawning from previous trajectories as well.

intensity $\nu(\cdot)$ and the cardinality distributions $p_{\mathrm{C}}(\cdot)$ are propagated in time by the CPHD recursion as

$$p_{\mathrm{C},\bar{k}|\bar{k}-1}(n) = \sum_{j=0}^{n} p_{\Gamma,\bar{k}}(n-j) \sum_{\ell=j}^{\infty} C_j^{\ell} \frac{\langle p_{\mathrm{S},\bar{k}}, \nu\rangle^j \langle 1 - p_{\mathrm{S},\bar{k}}, \nu\rangle^{\ell-j}}{\langle 1, \nu\rangle^{\ell}}, \tag{67}$$

$$\nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}) = \int p_{\mathrm{S},\bar{k}}(\boldsymbol{\xi})\, \varphi_{\bar{k}|\bar{k}-1}(\boldsymbol{x}|\boldsymbol{\xi})\, \nu_{\bar{k}-1}(\boldsymbol{\xi})\, d\boldsymbol{\xi} + \gamma(\boldsymbol{\xi}), \tag{68}$$

and

$$p_{\mathrm{C},\bar{k}|\bar{k}}(n) = \frac{\Upsilon_k^0[\nu_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}] p_{\mathrm{C},\bar{k}|\bar{k}-1}(n)}{\langle \Upsilon_k^0[\nu_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}], p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle}, \tag{69}$$

$$\begin{aligned} \nu_{\bar{k}|\bar{k}}(\boldsymbol{x}) = {} & \frac{\langle \Upsilon_{\bar{k}}^1[\nu_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}] p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle}{\langle \Upsilon_k^0[\nu_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}], p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle} \times [1 - p_{\mathrm{D},\bar{k}}(\boldsymbol{x})] \nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}) \\ & + \sum_{\boldsymbol{z}\in\boldsymbol{Z}_{\bar{k}}} \frac{\langle \Upsilon_{\bar{k}}^1[\nu_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}} \setminus \{z\}] p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle}{\langle \Upsilon_{\bar{k}}^0[\nu_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}], p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle} \times \psi_{\bar{k},\boldsymbol{z}}(\boldsymbol{x}) \nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}), \end{aligned} \tag{70}$$

where

$$\begin{aligned} \Upsilon_{\bar{k}}^u[\nu, \boldsymbol{Z}](n) = {} & \sum_{j=0}^{\min(\mathcal{C}(\boldsymbol{Z}),n)} (\mathcal{C}(\boldsymbol{Z})-j)! p_{K,\bar{k}}(\mathcal{C}(\boldsymbol{Z})-j) P_{j+u}^n \\ & \times \frac{\langle 1 - p_{\mathrm{D},\bar{k}}(\boldsymbol{x}), \nu\rangle^{n-(j+u)}}{\langle 1, \nu\rangle^n} e_j(\Xi_k(\nu, \boldsymbol{Z})), \end{aligned} \tag{71}$$

$$\psi_{\bar{k},\boldsymbol{z}}(\boldsymbol{x}) = \frac{\langle 1, \kappa_k\rangle}{\kappa_k(\boldsymbol{z})} h(\boldsymbol{z}|\boldsymbol{x}) p_{\mathrm{D},\bar{k}}(\boldsymbol{x}), \tag{72}$$

$$\Xi_k(\nu, \boldsymbol{Z})) = \{\langle \nu, \psi_{\bar{k},\boldsymbol{z}}\rangle : \boldsymbol{z} \in \boldsymbol{Z}\}, \tag{73}$$

with $\kappa_{\bar{k}}(\cdot)$ and $p_{K,\bar{k}}(\cdot)$ as the intensity of clutter and the distribution for the number of clutter arrivals at k, P_{j+u}^n the permutation coefficient, $e_j(\cdot)$ the elementary symmetric function, $p_{\Gamma,\bar{k}}$ as the cardinality distribution of the birth RFS and $\mathcal{C}(\cdot)$ the cardinality operator.

As for the PHD filter, the propagation of $\nu_{\bar{k}|\bar{k}}$ and $p_{\mathrm{C},\bar{k}|\bar{k}}$ in time would be too computationally demanding for any practical usage. To alleviate this problem in [136] proposed to approximate $\nu_{\bar{k}|\bar{k}}$ by a GM.

An interesting physical interpretation of both the PHD and the CPHD recursions in terms of the *bin-occupancy* filter is provided in [137] while the application of the filters to smoothing and prediction problems has been covered in [138].

Concluding, differently from the PHD filter where the cardinality of $\boldsymbol{X}_k$ is estimated using equation (66), with the CPHD recursion this can extracted directly from $p_{C,\bar{k}|\bar{k}}$ by means of MAP or MMSE estimators.

3.3 Performance metrics

The performance metric commonly used in active LT application and adopted here to asses the algorithm performance is the root mean square error (RMSE) computed over the location estimates $\hat{\boldsymbol{X}}$ and averaged over the number of realizations

$$\text{RMSE} \triangleq \tfrac{1}{\sqrt{N}} \cdot \|\boldsymbol{X} - \hat{\boldsymbol{X}}\|_{\text{F}}. \tag{74}$$

Since the RMSE measures the error in terms of distances, this metric makes the comparison of the results from different algorithms straightforward. It should be kept in mind that when dealing with NLOS conditions, RMSE measures may be heavily affected by outliers, making it a rather *coarse* performance measure when compared to the cumulative density function (cdf) of the error. However, since the vast majority of the scenarios considered in the present work deals with LOS observations only, it was decided to keep the RMSE as the metric that measures the performance in accuracy of the *active*-LT algorithms discussed in the sequel of this work.

Differently from active scenarios, the joint cardinality and targets' states estimate encountered in the passive MTT scenario makes this comparison more complicated. To overcome this difficulty, in this work the performance for the passive algorithms are compared using the optimal sub-pattern assignment (OSPA) metric recently proposed in [139].

Let $\bar{d}^{(c)}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \min(c, d(\boldsymbol{x}, \hat{\boldsymbol{x}}))$ be the distance between the two elements $(\boldsymbol{x}, \hat{\boldsymbol{x}}) \in \mathcal{X}$ with $c > 0$ as the cutoff value and Π_b the permutations for the set $\{1, \dots, b\}$ with $b \in \mathbb{N} = \{1, 2, \dots\}$, then the OSPA metric $\bar{d}_p^{(c)}$ is computed as

$$\bar{d}_p^{(c)}\left(\boldsymbol{X}, \hat{\boldsymbol{X}}\right) = \left(\frac{1}{n}\left(\min_{\pi \in \Pi_n} \sum_{i=1}^{m} d^{(c)}\left(\boldsymbol{x}_i, \hat{\boldsymbol{x}}_{\pi(i)}\right)^p + c^p(n-m)\right)\right)^{\frac{1}{n}}, \tag{75}$$

if $m \leqslant n$ and $\bar{d}_p^{(c)}\left(\boldsymbol{X}, \hat{\boldsymbol{X}}\right) = \bar{d}_p^{(c)}\left(\hat{\boldsymbol{X}}, \boldsymbol{X}\right)$ if $m \geqslant n$; moreover,

$$\bar{d}_\infty^{(c)}\left(\boldsymbol{X}, \hat{\boldsymbol{X}}\right) = \begin{cases} \min\limits_{\pi \in \Pi_n} \max\limits_{1 \leqslant i \leqslant n} d^{(c)}\left(\boldsymbol{x}_i, \hat{\boldsymbol{x}}_{\pi(i)}\right) & \text{if } \ m = n, \\ c & \text{otherwise.} \end{cases} \tag{76}$$

In the equation above the parameter p is used to weight the outliers differently, while c assigns a penalty proportional to the cutoff value to faulty assignments. Specifically, the lower the parameter c, the higher the relevance of tracking accuracy in the OSPA value, conversely, the higher c is, the more is the emphasis of cardinality error into the OSPA metric.

The other type of error that must be considered in passive MTT scenarios is the cardinality error which is here measured by

$$\mathrm{E}\left[\left|\mathcal{C}(\hat{\boldsymbol{X}}_k) - \mathcal{C}\left(\boldsymbol{X}_k\right)\right|\right],$$

where $|\cdot|$ denotes the absolute value operator.

As already mentioned above, one objective of this work is to investigate how to lower the computational load of current state of the art solutions for MTT-LT applications without affecting the localization accuracy. Since a comparison in terms of floating point operations would be too dependent on the specific machine and implementation used during the evaluation, it was decided to perform such a comparison using specific variables directly related to the state of the art solution we improve upon.

Specifically, for the subspace tracking technique proposed in Chapter 4, the computational time is measured in Section 4.3 in terms of the number of times the kernel matrix $\mathbf{K}$ is "swept" by plane rotations. Differently, in the *passive* case the computational complexity is measure in terms of number of *gated* measurements that are fed to the CPHD filter.[39]

3.4 Summary and discussions

The chapter focused on the different formulations of the multitarget LT problem in active and passive scenarios that will be used in the sequel of this work.

In particular relying on the STRAIN formulation of the active multitarget problem in dynamic scenario in Chapter 4 a scheme to track the eigenspace

[39] Notice that also a comparison in terms of computational time is offered in Table 2.

associated to the kernel matrix used in the C-MDS formulation of the problem is proposed. Chapter 5 deals with two limitations of the aforementioned technique, namely the data erasure problem over the EDM used to infer the targets location and the restriction of the approach to distance, or equivalently TOA observations.

The Bayesian approach instead is exploited in Chapter 6 to solve, by means of the RFS formulation, the MTT problem in passive scenarios. In this context it is shown that by integrating the S-CKF filter into the GM-CPHD solution it is possible to improve robustness of the resulting filter versus numerical problem. More importantly a novel gating scheme is proposed to limit the computational complexity of the approach as well as to improve the accuracy of the filter.

4 The active MTT Problem: complete scenarios

This chapter focuses on an efficient multitarget LT algorithm for the MTT problem in active scenarios using TOA-based range measurements between all the agents to be localized.

In so doing we use the C-MDS approach, that recently, due to its non-parametric formulation, has been shown an advantageous technique compared to more conventional methods especially when the targets' mobility cannot be simply and reliably modeled *a priori* [43–45, 129]. A problematic characteristic of the MDS method, however, is the need to continuously eigen-decompose the typically large Euclidean kernel constructed from the observations, which in turn calls for efficient eigen-decomposition algorithms capable of updating the eigenspace associated with an earlier kernel into the one of the latest observation [140–142],[143, pp. 431]. With regard to the latter, two important questions arise which motivate and the subject of the chapter:

- how large may the subspace distance of two subsequent kernels be, such that a subspace-tracking approach is still advantageous
- how to exploit the specific structure of such kernels so as to maintain complexity as low as possible.

With regard to the latter issue, in this chapter a statistically-optimized and truncated variation of the Jacobi algorithm, designed specifically to suit the structure of the dynamic MDS kernels under consideration is proposed. This modified Jacobi technique, which is founded on a statistical Gershgorin-theoretical eigen-spectrum analysis, mitigates the latter problem resting on the statistically ensured fact that only a fraction of the elements of dynamic MDS kernels need be swept by Givens rotations. Consequently, the computational cost associated with the eigen-decomposition of dynamic MDS kernels is brought down to a remarkably low complexity[40] of order $\mathcal{O}(\sqrt{N})$.

[40] Measured in terms of the number of plane rotation sweeps normalized to the subspace distance.

4.1 The C-MDS algorithm in dynamic scenarios

4.1.1 Static MDS kernels

Let $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ denote the matrix containing the set of cartesian coordinates for N points in an η-dimensional space, and let the dissimilarity of two points at locations $\mathbf{x}_i$ and $\mathbf{x}_j$ be given by their pairwise Euclidean distance defined in equation (10). Let $\mathbf{D}$ denote the squared EDM [40, pp.186] defined in equation (11). Next, consider the double-centered Euclidean kernel constructed in equation (18). In the C-MDS algorithm [144] the coordinate matrix $\mathbf{X}$ is recovered – up to rigid motion[41] – from the kernel[42] $\mathbf{K}$ by

$$\mathbf{X} = [\mathbf{V}]_{1:N,1:\eta} \cdot \mathbf{\Sigma} = [\mathbf{V}]_{1:N,1:\eta} \cdot [\mathbf{\Lambda}]^{\frac{1}{2}}_{1:\eta,1:\eta}, \tag{77}$$

where $(\mathbf{V}, \mathbf{\Lambda})$ is the eigen-pair of $\mathbf{K}$ and $\mathbf{\Sigma} \triangleq [\mathbf{\Lambda}]^{\odot\frac{1}{2}}_{1:\eta,1:\eta}$.

4.1.2 Dynamic scenarios

Let $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ and $\mathbf{X}_{\bar{T}} \in \mathbb{R}^{N\times\eta}$ denote two consecutive observations of a time variant real random process corresponding to the cartesian coordinates of objects to be tracked. Assume that $\mathbf{X}$ is perfectly known and, without loss of generality, represent $\mathbf{X}_{\bar{T}} \triangleq \mathbf{X} + \mathbf{Y}$, where the *dislocation* matrix $\mathbf{Y}$ is unknown, and consequently so is $\mathbf{X}_{\bar{T}}$.

Consider now the associated Gramian matrices

$$\mathbf{G} \triangleq \mathbf{X} \cdot \mathbf{X}^{\mathrm{T}}, \tag{78}$$

$$\mathbf{G}_{\bar{T}} \triangleq \mathbf{X}_{\bar{T}} \cdot \mathbf{X}^{\mathrm{T}}_{\bar{T}} = \mathbf{G} + \mathbf{Y} \cdot \mathbf{Y}^{\mathrm{T}} + \mathbf{X} \cdot \mathbf{Y}^{\mathrm{T}} + \mathbf{Y} \cdot \mathbf{X}^{\mathrm{T}}. \tag{79}$$

Let $(\mathbf{V}, \mathbf{\Lambda})$ denote the eigen-pair of $\mathbf{G}$ such that $\mathbf{G} = \mathbf{V} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^{\mathrm{T}}$ and consider the Gramian matrix $\mathbf{G}^{*}_{\bar{T}}$ obtained from $\mathbf{G}_{\bar{T}}$ by the eigenstructure-preserving similarity transformation

$$\mathbf{G}^{*}_{\bar{T}} \triangleq \mathbf{V}^{\mathrm{T}} \cdot \mathbf{G}_{\bar{T}} \cdot \mathbf{V} = \mathbf{\Lambda} + \mathbf{A} + \mathbf{B} + \mathbf{B}^{\mathrm{T}}, \tag{80}$$

[41] Rigid motion includes scaling, rotation and shift, which can be corrected by the Procrustes transformation detailed in Appendix 1.

[42] The kernel $\mathbf{K}$ obtained from equation (18) with $\mathbf{J}$ as in equation (19) is here referred to as the *static* MDS kernel in allusion to the fact that it is not constructed under the knowledge of the eigen-structure of previous data.

where $\mathbf{\Lambda} \triangleq \mathrm{diag}(\{\lambda_1, \cdots, \lambda_\eta, 0 \cdots, 0\})$, with $\lambda_1 \geqslant \cdots \geqslant \lambda_\eta$ and

$$\mathbf{A} \triangleq \mathbf{V}^{\mathrm{T}} \cdot \left(\mathbf{Y} \cdot \mathbf{Y}^{\mathrm{T}}\right) \cdot \mathbf{V}, \tag{81}$$

$$\mathbf{B} \triangleq \mathbf{V}^{\mathrm{T}} \cdot \mathbf{X} \cdot \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{V}. \tag{82}$$

Furthermore, let the *dominant* spaces of $\mathbf{G}$ and $\mathbf{G}_{\bar{T}}$ be defined as the η-larger-column partitions of their corresponding eigenvector matrices, respectively denoted by $[\mathbf{V}]_{1:N,1:\eta}$ and $[\mathbf{V}_{\bar{T}}]_{1:N,1:\eta}$, such that the subspace distance between $\mathbf{G}$ and $\mathbf{G}_{\bar{T}}$ can be defined by [143, pp.76]

$$\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}}) \triangleq \| \, [\mathbf{V}]_{1:N,1:\eta} \cdot [\mathbf{V}]^{\mathrm{T}}_{1:N,1:\eta} - [\mathbf{V}_{\bar{T}}]_{1:N,1:\eta} \cdot [\mathbf{V}_{\bar{T}}]^{\mathrm{T}}_{1:N,1:\eta} \, \|_{\mathrm{F}}, \tag{83}$$

where $\| \cdot \|_{\mathrm{F}}$ denotes Frobenius norm.

Clearly, if the subspace distance between $\mathbf{G}$ and $\mathbf{G}_{\bar{T}}$ is sufficiently small, equation (80) *quasi*-diagonalizes[43] $\mathbf{G}_{\bar{T}}$ to an extent inversely proportional to $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$.

Several problems in signal processing – including the multitarget tracking problem of interest here – amount to obtaining estimates of the matrices $\{\mathbf{X}, \mathbf{X}_{\bar{T}}, \mathbf{X}_{2\bar{T}} \cdots\}$ given a sequence of Gramian matrices $\{\mathbf{G}, \mathbf{G}_{\bar{T}}, \mathbf{G}_{2\bar{T}} \cdots\}$ periodically "observed" (*i.e.* either acquired directly or constructed out of other measurable quantities).

Of relevance to such applications is the well-known fact that under the above conditions, sequential eigen-decomposition algorithms (*e.g.* [143, pp.431]) can exploit the similarity of $\mathbf{G}$ and $\mathbf{G}_{\bar{T}}$ – in terms of a sufficiently small subspace distance $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ – to obtain the eigen-pairs of the sequence of matrices $\{\mathbf{G}_{\bar{T}}, \mathbf{G}_{2\bar{T}}, \cdots\}$ successively.

What is not as well-known, however, is *how small* must the subspace distance $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ be for such eigen-adaptation approach to be advantageous. In other words, *how large* may $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ be for the eigen-pair $(\mathbf{V}, \mathbf{\Lambda})$ to still contain sufficient information on the eigen-pair $(\mathbf{V}_{\bar{T}}, \mathbf{\Lambda}_{\bar{T}})$ of $\mathbf{G}_{\bar{T}}$ to justify the adaptation of $(\mathbf{V}, \mathbf{\Lambda})$ into $(\mathbf{V}_{\bar{T}}, \mathbf{\Lambda}_{\bar{T}})$, as opposed to the direct computation of $(\mathbf{V}_{\bar{T}}, \mathbf{\Lambda}_{\bar{T}})$? While an answer to this question may never be found generally (that is, for arbitrary matrices), the answer for the particular case of the Gramian matrices specified above is one of the contributions of this chapter. In particular, it

[43] The term *quasi*-diagonal refers to an off-diagonal square-norm sense clarified in Section 4.3.2.

is shown that in the most relevant case[44] where $\mathbf{Y}$'s are *random* zero-mean Gaussian matrices with independent and identically distributed (i.i.d.) elements, even for substantially large $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$, the dynamic kernel $\mathbf{G}^*_{\bar{T}}$ can (with high probability) be diagonalized at a fraction of the cost typically associated with the eigen-decomposition[45].

Notice also that it is well-known to those familiar with random matrix theory that, in general, one "*cannot find the eigenvalues of the sums of random matrices from the eigenvalues of the individual matrices*" [145, pp. 19]. In order to obtain our result, however, in Section 4.2 we will make use of the elegant theory of Gershgorin spectrum bounds.

4.1.3 Dynamic MDS kernels

In coherence with the notation of preceding subsections, denote the squared Euclidean distance matrix associated to $\mathbf{X}_{\bar{T}} = \mathbf{X} + \mathbf{Y}$ by $\mathbf{D}_{\bar{T}}$. Next, consider the alternative *dynamic* MDS kernel[46] $\mathbf{K}^*$ constructed from $\mathbf{D}_{\bar{T}}$ as defined in equation (18), but with $\mathbf{J}$ replaced by $\mathbf{J}^*$

$$\mathbf{J}^* = \mathbf{V} - \tfrac{1}{N} \cdot \mathbf{1}_N \cdot \left[\mathbf{1}_N^{\mathrm{T}} \cdot \mathbf{v}_1 \,\vdots\, \cdots \,\vdots\, \mathbf{1}_N^{\mathrm{T}} \cdot \mathbf{v}_N \right], \tag{84}$$

where $\mathbf{v}_i$ is the i-th column of $\mathbf{V}$, the eigenvector of the preceding kernel (see equation (87)).

Notice that the structure of $\mathbf{J}^*$ is very similar to that of $\mathbf{J}$ and therefore the dynamic kernel $\mathbf{K}^*$ is constructed essentially as efficiently as the static kernel $\mathbf{K}$. Furthermore, observe that the conventional MDS algorithm operates with static kernels and consequently its direct utilization in tracking applications amount to a re-localization technique. In contrast, the MDS method over dynamic kernels constructed using equation (84) has a truly tracking nature, as it maps a trajectory tracking problem onto a subspace tracking problem.

Now, consider the following result.

[44] Under the absence of *a priori* mobility information, zero-mean Gaussianity amounts to the weakest possible assumption on the directions targets may move to, *i.e.*, maximally entropic dislocation likelihood with equiprobable directions.

[45] The complexity required by the Jacobi algorithm to eigen-decompose a symmetric matrix employing a standard sequential sweeping strategy is used as reference.

[46] The slight abuse of not including the subscript $_{\bar{T}}$ in the notation of $\mathbf{K}$ will prove convenient shortly.

Lemma 1 (Equivalence of $\mathbf{K}^*$ and $\mathbf{G}^*_{\bar{T}}$)**.** *Let the coordinate matrices* $\mathbf{X}_{\bar{T}} \in \mathbb{R}^{N\times\eta}$ *be outcomes of an uncorrelated zero-mean point process. Then for* $N \to \infty$ *the kernel matrix* $\mathbf{K}^*$ *converges in probability to* $\mathbf{G}^*_{\bar{T}}$, i.e

$$\lim_{N\to\infty} \Pr\{\mathbf{K}^* - \mathbf{G}^*_{\bar{T}}\} = 0. \tag{85}$$

Proof. : The proof is trivial and thus given in a concise form. First notice that $\mathbf{J}^* = \mathbf{J}\cdot\mathbf{V}$ and therefore $\mathbf{K}^* = \mathbf{V}^{\mathrm{T}}\cdot\mathbf{K}\cdot\mathbf{V}$. Now, $\mathbf{K}$ is known as the *double-centered* Euclidean kernel due to the equivalence [40, pp. 195-196] $\mathbf{K}=(\mathbf{X}_{\bar{T}}-\bar{\mathbf{X}}_{\bar{T}})\cdot(\mathbf{X}_{\bar{T}}-\bar{\mathbf{X}}_{\bar{T}})^{\mathrm{T}}$, where $\bar{\mathbf{X}}_{\bar{T}}$ denotes the sample mean of $\mathbf{X}_{\bar{T}}$. But by assumption $\mathrm{E}[\mathbf{X}] = \mathbf{0}_{[N\times\eta]}$ – where $\mathrm{E}[\cdot]$ denotes expectation[47] – such that $\lim_{N\to\infty}\bar{\mathbf{X}}_{\bar{T}} = \mathbf{0}_{[N\times\eta]}$, which concludes the proof. □

Before we conclude this subsection, let us point out that although MDS kernels in principle require complete Euclidean distance matrices $\mathbf{D}_{\bar{T}}$ to be constructed, as Chapter 5 shows, this issue can be overcome using a technique referred to as *Nyström approximation*[48] [146–148].

4.2 Gershgorin spectral analysis of random Gramian matrices

Gershgorin's elegant and powerful theorem is most often utilized to study the properties of deterministic matrices [149]. For instance, the radius of the Gershgorin discs of a given matrix can be used to verify how close the matrix is to its diagonal form [150]. Likewise, the positive definiteness of a matrix can be studied by simply investigating if its Gershgorin discs all lie on the right complex semi-plane [151]. Here, however, due to the randomness of the unknown dislocation $\mathbf{Y}$ and the arbitrariness of the location matrix $\mathbf{X}$, which together determines $\mathbf{X}_{\bar{T}}$, the characterization of the *quasi*-diagonality of $\mathbf{G}^*_{\bar{T}}$ can only be performed in a statistical sense.

[47] In this work, expectations with respect to different domains are considered, depending on corresponding contexts. For instance, here we refer to expectation with respect to the N number of points in the space. In the context of Section 4.2, however, expectation is considered with respect to different outcomes of a random variate.

[48] In the case of Euclidean kernels, the Nyström "approximation" is actually an exact completion if the entries of the required minors are error-free.

For convenience, let us briefly review and contextualize the theory of Gershgorin bounds.

Let $\mathbf{K} \in \mathbb{R}^{N\times N}$ be a square matrix decomposed into the sum $\mathbf{K} = \boldsymbol{\Upsilon} + \boldsymbol{\Omega}$, where $\boldsymbol{\Upsilon}$ is diagonal and $\boldsymbol{\Omega}$ is a matrix whose diagonal entries are all zero (*i.e.*, hollow). Let also $\upsilon_i = [\boldsymbol{\Upsilon}]_{ii}$, $\omega_{ij} = [\boldsymbol{\Omega}]_{ij}$ and $\bar{\omega}_i \triangleq \sum_{j=1}^{N} |\omega_{ij}|$, then:

- The i-th **Gershgorin disc** associated to the i-th row of $\mathbf{K}$ is defined as the interval $\mathbb{D}_i \triangleq [\upsilon_i - \bar{\omega}_i, \upsilon_i + \bar{\omega}_i] \in \mathbb{R}$. Accordingly, the quantities υ_i and $\bar{\omega}_i$ are respectively referred to as the center and the radius of $\mathbb{D}_i$.
- Let $\mathcal{I} \triangleq \{1, \cdots, N\}$ be the set of indexes associated with the Gershgorin discs and define its powerset $\mathcal{P}(\mathcal{I}) \triangleq \{\emptyset, \{1\}, \cdots, \{N\}, \{1,2\}, \{1,3\}, \cdots, \mathcal{I}\} \equiv \{\emptyset, \mathcal{P}_1, \cdots, \mathcal{P}_{2^N-1}\}$. The **Gershgorin group** $\mathbb{G}_k$ is defined as

$$\mathbb{G}_k \triangleq \bigcup_{i\in\mathcal{P}_k} \mathbb{D}_i. \tag{86}$$

- A pair of Gershgorin groups $\mathbb{G}_k$ and $\mathbb{G}_q$ is said to be **disjoint** if and only if (iff) $\mathbb{G}_k \cap \mathbb{G}_q = \emptyset$. Otherwise, $\mathbb{G}_k$ and $\mathbb{G}_q$ are said to be **joint**.
- The **cardinality** of a Gershgorin group $\mathbb{G}_k$ is denoted $\mathcal{C}(\mathbb{G}_k)$ and defined as the cardinality of the **associated set of indexes** $\mathcal{P}_k$, *i.e.*, $\mathcal{C}(\mathbb{G}_k) \triangleq \mathcal{C}(\mathcal{P}_k)$.
- Consider any[49] bijection between $\{\mathbb{G}_1, \cdots, \mathbb{G}_{2^N-1}\}$ and the set given by $\{\tilde{\mathbb{G}}_1, \cdots, \tilde{\mathbb{G}}_Q, \tilde{\mathbb{G}}_{Q+1}, \cdots, \tilde{\mathbb{G}}_{2^N-1}\}$ and satisfying the conditions:

  1. All Gershgorin Groups in $\mathcal{G}_Q \triangleq \{\tilde{\mathbb{G}}_1, \cdots, \tilde{\mathbb{G}}_Q\}$ are mutually disjoin
  2. Q is the largest possible number
  3. $\mathbb{D}_i \subseteq \mathcal{G}_Q$, $\forall\, i \in \mathcal{I}$.

  Then, each disjoint Gershgorin Group $\tilde{\mathbb{G}}_q \in \mathcal{G}_Q$ is referred to as a **Gershgorin superdisc**.
- Out of the Q Gershgorin superdiscs of a matrix $\mathbf{K} \in \mathbb{R}^{N\times N}$ of $\text{rank}(\mathbf{K}) < N$, the one that contains the origin is referred to as the **Gershgorin null superdisc** is denoted $\mathbb{S}_0$.

We are now ready to introduce the Gershgorin Theorem in a concise form convenient to the needs hereafter.

[49] The order of $\tilde{\mathbb{G}}_q$ is arbitrary, and the number $Q \in [1, N]$ is uniquely determined by the three conditions (see Corollary 1).

Theorem 1 (Gershgorin Theorem)**.** *Let* $\{\lambda_1, \cdots, \lambda_N\}$ *be the* N *eigenvalues of* $\mathbf{K} \in \mathbb{R}^{N\times N}$*. Then*

$$\mathcal{C}(\tilde{\mathbb{G}}_q) = \sum_{i=1}^{N} \mathcal{F}_{\tilde{\mathbb{G}}_q}(\lambda_i), \forall\, q \leqslant Q,$$

where $\mathcal{F}_{\tilde{\mathbb{G}}_q}$ denotes the indicator function for the superdisc $\tilde{\mathbb{G}}_q$.

Proof. See [149, 150]. □

In plain words the Gershgorin Theorem establishes that each superdisc $\mathbb{S}_k$ contains *exactly* as many eigenvalues of $\mathbf{K}$ as the number of discs that compose it. And since the discs are constructed directly form the elements of $\mathbf{K}$, the theorem draws a direct relationship between the eigen-spectrum of a matrix and its elements.

To conclude, notice that an important corollary of Lemma 1 is that the analysis performed in the following subsections over $\mathbf{G}^*_{\bar{T}}$ applies straightforwardly to the dynamic kernels $\mathbf{K}^*$ constructed as specified in Section 4.1.3.[50]

4.2.1 Structure of $\mathbf{G}^*_{\bar{T}}$

To gain insight, first recall from Section 4.1.2 that the Gramian matrices under consideration are η-ranked, and consider the matrix $\mathbf{V}^{\mathrm{T}} \cdot \mathbf{G} \cdot \mathbf{V} = \mathbf{\Lambda}$. Obviously all the Gershgorin discs or superdiscs of the diagonal matrix $\mathbf{\Lambda}$ are deteriorated to (at most) $\eta + 1$ points, specifically, the multiset of the non-zero eigenvalues $\{\lambda_1, \cdots, \lambda_\eta\}$ and the null eigenvalue 0.

Next, consider the similarity transformation of equation (80). By force of the Wielandt-Hoffman Theorem [143, pp.395], the Gershgorin discs of $\mathbf{G}^*_{\bar{T}}$ are enlarged compared to those (points) of $\mathbf{\Lambda}$, and their centers are likely to be shifted away from $\{\lambda_1, \cdots, \lambda_\eta, 0\}$. And by force of the Gershgorin Theorem and the fact that $\mathrm{rank}(\mathbf{G}^*_{\bar{T}}) = \eta$, it is certain that the null superdisc $\mathbb{S}^*_0$ of $\mathbf{G}^*_{\bar{T}} \in \mathbb{R}^{N\times N}$ contains *at least* $N - \eta$ null eigenvalues. What cannot be known *a priori*, however, is if $\mathbb{S}^*_0$ is expanded enough by the terms $\mathbf{A} + \mathbf{B} + \mathbf{B}^{\mathrm{T}}$ to include one or more non-zero eigenvalue of $\mathbf{G}^*_{\bar{T}}$, and if $\mathbb{S}^*_0$ is still associated to its *bottom* $N - \eta$ rows.

[50] As shown in [40, pp. 194] the equivalence between $\mathbf{K}^*$ and $\mathbf{G}^*_{\bar{T}}$ can be established independently of N using the full-rank skinny Schoenberg auxiliary matrix to center the EDM $\mathbf{D}$ to the origin.

The occurrence of either of these phenomena indicates that $\mathbf{G}_{\bar{T}}$ is not brought sufficiently close to a *quasi*-diagonal form (orderly, and otherwise unaffectedly) by the similarity transformation of equation (80). In order to quantify the probability of the latter, a statistical analysis of the Gershgorin discs of $\mathbf{G}^*_{\bar{T}}$ is required. This, in turn, calls for the statistical characterization of the elements of $\mathbf{G}^*_{\bar{T}}$. From equation (81), it is clear that if the increment matrix $\mathbf{Y}$ is statistically dense so is the matrix $\mathbf{A}$. As a consequence, the structure of $\mathbf{B}$ can be better understood by expanding equation (82) as follows.

First, consider the singular value decomposition (SVD) of $\mathbf{X}$

$$\mathbf{X} = \mathbf{V} \cdot \mathbf{\Sigma} \cdot \mathbf{U}^{\mathrm{T}}, \tag{87}$$

where we used the facts that the left singular-vector and singular-value matrices of $\mathbf{X}$ are identical to $\mathbf{V}$ and $[\mathbf{\Lambda}]^{\odot\frac{1}{2}}_{1:\eta,1:\eta}$, respectively (see Section 4.3).

From equations (82) and (87) we readily obtain

$$\mathbf{B} = \mathbf{\Sigma} \cdot \mathbf{U}^{\mathrm{T}} \cdot \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{V}, \tag{88}$$

which, due to the structure of $\mathbf{\Sigma}$, implies that the elements of $\mathbf{B}$ below the η-th row are zero.

Next, let us rewrite $\mathbf{G}^*_{\bar{T}}$ in the form

$$\mathbf{G}^*_{\bar{T}} = \mathbf{\Upsilon} + \mathbf{\Omega}, \tag{89}$$

where $\mathbf{\Upsilon}$ is the diagonal matrix

$$\mathbf{\Upsilon} = \left[\begin{array}{c:c} \mathrm{diag}(\boldsymbol{\varrho})_{\eta\times\eta} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathrm{diag}(\boldsymbol{\varepsilon})_{(N-\eta)\times(N-\eta)} \end{array}\right], \tag{90}$$

in which $\mathrm{diag}(\boldsymbol{\varrho})$ and $\mathrm{diag}(\boldsymbol{\varepsilon})$ denote the diagonal matrix forms of the vectors $\boldsymbol{\varrho}$ and $\boldsymbol{\varepsilon}$, respectively, and $\mathbf{\Omega}$ is a hollow symmetric matrix with structure

$$\mathbf{\Omega} = \left[\begin{array}{c:c} \mathbf{\Phi}_{\eta\times\eta} & \mathbf{\Theta}_{\eta\times(N-\eta)} \\ \hdashline \mathbf{\Theta}^{\mathrm{T}}_{\eta\times(N-\eta)} & \mathbf{\Psi}_{(N-\eta)\times(N-\eta)} \end{array}\right]. \tag{91}$$

The partitions shown in equations (90) and (91) are there to indicate the fact that, given a random $\mathbf{Y}$, the random matrices $\mathbf{\Phi}$, $\mathbf{\Psi}$, $\mathbf{\Theta}$, $\mathrm{diag}(\boldsymbol{\varrho})$ and $\mathrm{diag}(\boldsymbol{\varepsilon})$ all

follow distinct statistics, which will be described shortly. For now, suffice it to convene that the non-zero elements of these matrices be referred to as ϕ-type, ψ-type, θ-type, ϱ-type and ε-type random variates, respectively. In the sequel, all these random variates are characterized in their relation with the Gershgorin center (υ_i) and radii ($\bar{\omega}_i$).

4.2.2 Centers of Gershgorin discs of $\mathbf{G}^*_{\bar{T}}$

From Subsection 4.2.1 it follows that the centers of the Gershgorin discs of $\mathbf{G}^*_{\bar{T}}$ are determined by ϱ-type and ε-type random variates, about which the following holds.

Lemma 2 (Statistics of ϱ_i)**.** *Let the elements of the matrix* $\mathbf{Y}$ *be independent and identically distributed (iid) zero-mean Gaussian variates with variance* σ_y^2*, and* $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ *be an arbitrary matrix with SVD as in equation* (87)*. Consider furthermore the matrix* $\mathbf{G}^*_{\bar{T}}$ *in the form given in equation* (80) *and structured as in equation* (89) *and let* $r \sim p(r)$ *indicates that* r *has distribution* $p(r)$ *and* $p_{\text{Gauss}}(r;\mu,\sigma)$ *is the Gaussian probability density function (pdf) with mean* μ *and standard deviation* σ

$$p_{\text{Gauss}}(r;\mu,\sigma) \triangleq \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(\frac{-(r-\mu)^2}{2\sigma^2}\right). \tag{92}$$

Then

$$\varrho_i \geqslant \rho_i \sim p_{\text{Gauss}}\left(r;\lambda_i, 2\sigma_y\sqrt{\lambda_i}\right). \tag{93}$$

Proof. Referring to equations (80), (81) and (88), it is clear that $\varrho_i = \lambda_i + a_{ii} + 2b_{ii}$ with $1 \leqslant i \leqslant \eta$. Next, notice[51] that $b_{ii} \sim p_{\text{Gauss}}\left(r;0,\sigma_y\sqrt{\lambda_i}\right)$, while $a_{ii} \sim p_{\text{Gamma}}(r;\eta/2,2\sigma_y^2)$, where

$$p_{\text{Gamma}}(r;\gamma,\delta) \triangleq \frac{r^{\gamma-1}}{\Gamma(\gamma)\cdot\delta^\gamma} \cdot \exp\left(-\frac{r}{\delta}\right). \tag{94}$$

Next, define $\rho_i \triangleq (\lambda_i + 2b_{ii})$, such that $\varrho_i \geqslant \rho_i$ since a_{ii} is non-negative, and finally recognize that due to the distribution of b_{ii}, $\rho_i \sim p_{\text{Gauss}}\left(r;\lambda_i,2\sigma_y\sqrt{\lambda_i}\right)$. □

The advantage of using ρ_i to lower bound ϱ_i in the context of our analysis will become clear soon. Let us also point out that while the relation $\rho_i \leqslant \varrho_i$

[51] See equations (81) and (82).

holds with probability 1, a similar but more general notion of bounding a random variable by another, with a certain probability, will be useful in the sequel. Thus the following definition.

Definition 1 (Upper/Lower-bound in Probability)**.** *Let y and x be random variates with arbitrary distributions. The notation $y \underset{p}{\geqslant} x$ indicates that y* ***is larger than*** *x with probability p, namely $\Pr\{y \geqslant x\} = p$. Likewise, $y \underset{p}{\leqslant} x$ indicates that y* ***is not larger than*** *x with probability p. It is said, respectively, that y is an upper/lower bound of x in probability p.*

Lemma 3 (Statistics of ε_i)**.** *Let the elements of the matrix $\mathbf{Y}$ be independent and identically distributed (iid) zero-mean Gaussian variates with variance σ_y^2, and $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ be an arbitrary matrix with SVD as in equation* (87)*. Consider furthermore the matrix $\mathbf{G}_{\bar{T}}^*$ in the form given in equation* (80) *and structured as in equation* (89)*. Then*

$$\varepsilon_{i_p} \underset{\epsilon}{\leqslant} \sim p_{\text{Rayleigh}}(r; R_\epsilon), \tag{95}$$

with bounding probability

$$p = \left(\frac{R_\epsilon}{16\sigma_y^4}\right)^{\frac{\eta}{4}} \cdot U\left(\frac{\eta}{4}, \frac{1}{2}, \frac{R_\epsilon}{16\sigma_y^4}\right), \tag{96}$$

where $U(\nu, \kappa; r)$ is the hypergeometric function known as the Gordon, Tricomi or Kummer's function of the second kind and $p_{\text{Rayleigh}}(r; R)$ is the Rayleigh pdf

$$p_{\text{Rayleigh}}(r; R) \triangleq \frac{2r}{R} \cdot \exp\left(\frac{-r^2}{R}\right). \tag{97}$$

Proof. See Appendix 2. □

Notice that equation (96) is in the form $p = r^\nu \cdot U(\nu, 1/2, r)$, which admits the asymptotic expression [152, pp.190] (for large r) $p \approx {}_2F_0(\nu, \nu + 1/2; ; -1/r)$. It follows therefore that $\lim_{r\to\infty} p = {}_2F_0(\nu, \nu + 1/2; ; 0) = 1$. In other words, as expected, the parameter R_ϵ can be chosen to ensure that the inequality $\varepsilon \underset{p}{\leqslant} \epsilon$ holds with any desired probability p.

Furthermore, $\mathrm{E}[\epsilon^2]/\mathrm{E}[\varepsilon^2] = R_\epsilon/(\eta^2 + 2\eta)\sigma_y^4$, such that equation (96) can be cast as a function of the square root of the second moment ratio, or else the *root mean square ratio* (rms-ratio) of ϵ and ε. In the appropriate context, this figure of merit is also known as the "power ratio" or signal-to-noise-ratio (SNR), and is

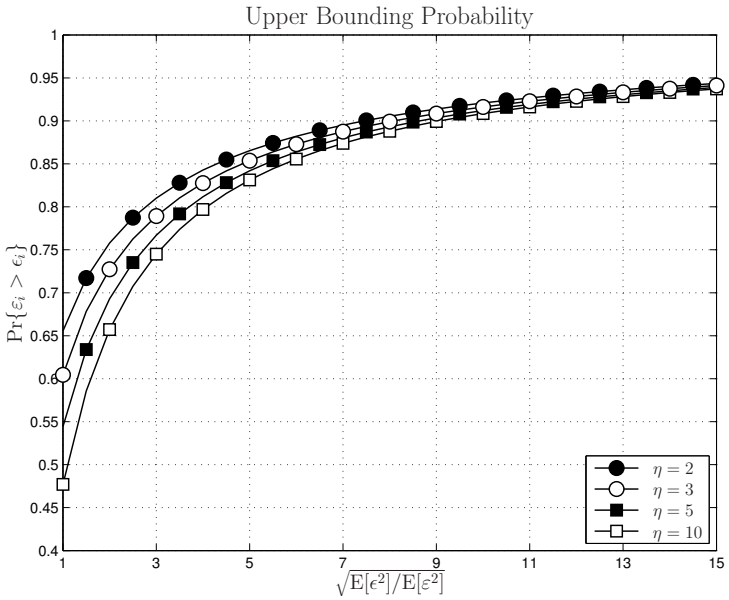


Fig 11. Upper bounding probability of ε as a function of the rms-ratio of ϵ and ε (see proof of Lemma 3). Revised from [12].

widely used as a measure of how larger a random variate is compared to another. Figure 11 shows plots of equation (246) as a function of the rms-ratio of ϵ and ε, which indicate that the Gamma-distributed variate ε is upper-bounded with high probability by Rayleigh variates ϵ with only slightly larger "power." For example, $p = 0.85$ if $\mathrm{E}[\epsilon^2] \approx 25\mathrm{E}[\varepsilon^2]$ for $\eta = 3$ and $\mathrm{E}[\epsilon^2] \approx 36\mathrm{E}[\varepsilon^2]$ for $\eta = 10$.

To conclude this subsection, let us point out that the approach of lower- and upper-bounding the variates ϱ and ε, respectively, leads to conservative estimates of the probability that $\mathbf{G}_{\bar{T}}$ is *quasi*-diagonalized by equation (80), therefore only strengthening the results hereafter.

4.2.3 Radii of Gershgorin discs of $\mathrm{G}^*_{\bar{T}}$

Referring to equation (91), recall that the evaluation of the radii of the Gershgorin discs of $\mathbf{G}^*_{\bar{T}}$ requires the statistics of the row-wise sum of the absolute values of ϕ-type and θ-type variates for the upper η rows, and of θ-type and ψ-type variates for the bottom $N - \eta$ rows. Specifically,

$$\bar{\omega}_i \triangleq \begin{cases} \sum\limits_{\substack{j=1 \\ j\neq i}}^{\eta} |\phi_{ij}| + \sum\limits_{j>\eta}^{N} |\theta_{ij}| \quad \text{for } 1 \leqslant i \leqslant \eta, & \text{(98a)} \\ \sum\limits_{j=1}^{\eta} |\theta_{ji}| + \sum\limits_{\substack{j>\eta \\ j\neq i}}^{N} |\psi_{ij}| \quad \text{for } i > \eta. & \text{(98b)} \end{cases}$$

In order to better understand the statistics of the ϕ-type, θ-type and ψ-type variates, let us first rewrite equations (81) and (88) as

$$\mathbf{A} = \mathbf{W} \cdot \mathbf{W}^{\mathrm{T}}, \tag{99}$$

$$\mathbf{B} = \mathbf{C} \cdot \mathbf{W}^{\mathrm{T}}, \tag{100}$$

where $(\mathbf{W}, \mathbf{C}) \in \mathbb{R}^{N\times\eta}$, but where c_{ij}'s are zero for $i > \eta$ and *constant* for $1 \leqslant (i,j) \leqslant \eta$, while $w_{ij} \sim p_{\mathrm{Gauss}}(r; 0, \sigma_y)$.

Thus we may write

$$a_{ij} = \sum_{k=1}^{\eta} w_{ik} \cdot w_{jk}, \quad \forall (i,j) \leqslant N, i \neq j, \tag{101}$$

$$b_{ij} = \begin{cases} \sum\limits_{k=1}^{\eta} c_{ik} \cdot w_{jk} & \text{for } 1 \leqslant i \leqslant \eta, \\ 0 & \text{otherwise.} \end{cases} \tag{102}$$

Clearly, for $i < j$, a_{ij}'s are i.i.d. random variates, moreover $a_{ij} = a_{ji}$.[52] Furthermore, it can be seen from equation (102) that the elements of the j-th column of $\mathbf{B}$ are different linear combinations of the same random variates $\{w_{j1}, \cdots, w_{j\eta}\}$. Due to the orthonormality of the matrix $\mathbf{U}$, the constant vectors $\mathbf{c}_i \triangleq [c_{i1}, \cdots, c_{i\eta}]$ are orthogonal with norm equal to λ_i (see equation (87)). Therefore, the rightmost summations in equation (98) can be written as

$$\sum_{\substack{j>\eta \\ j\neq i}}^{N} |\psi_{ij}| = \sum_{\substack{j>\eta \\ j\neq i}}^{N} \left| \sum_{k=1}^{\eta} w_{ik} \cdot w_{jk} \right|, \tag{103}$$

$$\sum_{j>\eta}^{N} |\theta_{ij}| = \sum_{j>\eta}^{N} \left| \sum_{k=1}^{\eta} (c_{ik} + w_{ik}) \cdot w_{jk} \right|, \tag{104}$$

where w_{ik} and $c_{ik} + w_{ik}$ are constant with respect to j,thus the following result.

[52] As equation (99) shows, the matrix $\mathbf{A}$ is Gram and therefore symmetric.

Lemma 4. *Consider random variates in the form*

$$z_i = \sum_{j=1}^{M} \left| \sum_{k=1}^{\eta} h_{ik} \cdot w_{jk} \right|, \tag{105}$$

*where w_{jk} are real-valued zero-mean Gaussian variates with variances σ^2 and distinct i.i.d. outcomes for each k and j, while h_{ik}'s are also Gaussian variates with variances σ^2, with an **arbitrary** mean, and i.i.d. outcomes with respect to i and k.*

Then, the asymptotic distributions of z_i (for large M) are

$$z_i \sim \begin{cases} p_{\mathrm{SC}\chi}(r;\eta,\alpha) \triangleq \dfrac{2\, r^{\eta-1}}{\alpha^{\eta/2} \cdot \Gamma(\eta/2)} \cdot \exp\left(\dfrac{-r^2}{\alpha}\right) & \text{if } \mathrm{E}[h_{ik}] = 0 \,\forall\, k, \\ p_{\mathrm{SNC}\chi}(r;\eta,\beta_i,\alpha) \triangleq \dfrac{2\, r^{\eta/2}}{\alpha \cdot \beta_i^{\eta/2-1}} \cdot \exp\left(-\dfrac{r^2+\beta_i^2}{\alpha}\right) \cdot I_{\eta/2-1}\left(\dfrac{2\beta_i r}{\alpha}\right) & \text{otherwise,} \end{cases} \tag{106}$$

where $p_{\mathrm{SC}\chi}(r;\eta,\alpha)$ and $p_{\mathrm{NSC}\chi}(r;\eta,\beta_i,\alpha)$ denote the central and non-central scaled χ distributions respectively and

$$\alpha \triangleq \frac{4M^2\sigma^4}{\pi}, \tag{107}$$

$$\beta_i \triangleq \sqrt{\frac{\alpha}{2} \cdot \sum_{k=1}^{\eta} \frac{\mathrm{E}[h_{ik}]^2}{\sigma^2}}. \tag{108}$$

Proof. See Appendix 3. □

Corollary 1 (Statistics of $\sum|\theta_{ij}|$ and $\sum|\psi_{ij}|$)**.** *Let the elements of the matrix $\mathbf{Y}$ be independent and identically distributed (iid) zero-mean Gaussian variates with variance σ_y^2, and $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ an arbitrary matrix with SVD as in equation* (87)*. Consider furthermore the matrix $\mathbf{G}_{\bar{T}}^{*}$ in the form given in equation* (80) *and structured as in equation* (89)*. Then, the asymptotic distributions of $\sum_{j>\eta}^{N} |\theta_{ij}|$ and $\sum_{\substack{j>\eta \\ j\neq i}}^{N} |\psi_{ij}|$ (for large N) are*

$$\sum_{j>\eta}^{N} |\theta_{ij}| \sim p_{\mathrm{SNC}\chi}\Big(r;\eta,(N-\eta)\sigma_y\sqrt{2\lambda_i/\pi},4(N-\eta)^2\sigma_y^4/\pi\Big), \tag{109}$$

$$\sum_{\substack{j>\eta \\ j\neq i}}^{N} |\psi_{ij}| \sim p_{\mathrm{SC}\chi}(r;\eta,4(N-\eta-1)^2\sigma_y^4/\pi). \tag{110}$$

Proof. For the case of the upper η rows, notice that h_{ik} in Lemma 4 are the constants c_{ik} such that $\mathrm{E}\Big[\big(\sum_{k=1}^{\eta} h_{ik}^2\big)^{n/2}\Big] = \lambda_i^{n/2}$. Furthermore, recognize that $M = (N - \eta)$, which immediately yields equation (109). □

In turn, for the lower $N - \eta$ rows, $\mathrm{E}[h_{ik}] = 0$ and $M = (N - \eta - 1)$, leading to equation (110).

As for the remaining terms in equation (98) the following result applies.

Lemma 5 (Statistics of $|\phi_{ij}|$ and $|\theta_{ji}|$)**.** *Let the elements of the matrix* $\mathbf{Y}$ *be independent and identically distributed (iid) zero-mean Gaussian variates with variance* σ_y^2*, and* $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ *an arbitrary matrix with SVD as in equation* (87)*. Consider furthermore the matrix* $\mathbf{G}_{\bar{T}}^*$ *in the form given in equation* (80) *and structured as in equation* (89)*. Then, the asymptotic distributions of* $|\phi_{ij}|$ *and* $|\theta_{ji}|$ *(for large* η*) are*

$$|\phi_{ij}| \sim p_{\text{Nakagami}}\left(r; \tfrac{1}{2}, (\eta\sigma_y^2 + \lambda_i + \lambda_j)\sigma_y^2\right), \tag{111}$$

$$|\theta_{ji}| \sim p_{\text{Nakagami}}\left(r; \tfrac{1}{2}, (\eta\sigma_y^2 + \lambda_j)\sigma_y^2\right). \tag{112}$$

Proof. See Appendix 4. □

Despite being asymptotic, the above results are found to accurately characterize the statistics of the Gershgorin radii $\bar{\omega}_i$ also for reasonably small η and N.

4.2.4 Cardinality of the null superdisc

To return to the discussion that initiated this section, the results of the preceding Subsections can now be employed to assess the effectiveness of equation (80) in *quasi*-diagonalizing $\mathbf{G}_{\bar{T}}$ while preserving its row structure. Equivalently the above results can be used to estimate the likelihood that the cardinality of the Gershgorin null superdisc of $\mathbf{G}_{\bar{T}}^*$ surpasses the limit $N - \eta$. With respect to the latter, the following result can be stated.

Theorem 2 (Cardinality of Null Superdisc of $\mathbf{G}_{\bar{T}}^*$)**.** *Let the elements of the matrix* $\mathbf{Y}$ *be independent and identically distributed (iid) zero-mean Gaussian variates with variance* σ_y^2*, and* $\mathbf{X} \in \mathbb{R}^{N\times\eta}$ *be an arbitrary matrix with SVD as in equation* (87)*. Let the descending list* $\{\lambda_1, \cdots, \lambda_\ell, \cdots, \lambda_\eta\}$ *of non-zero*

eigenvalues of the matrix $\mathbf{G}_{\bar{T}}$ *be known, and consider the matrix* $\mathbf{G}^*_{\bar{T}}$ *as given in equation* (80) *with structure as in equation* (89)*, and recall that* $\mathbb{S}^*_0$ *denotes the null superdisc of* $\mathbf{G}^*_{\bar{T}}$*. Then,*

$$\Pr\left\{\mathcal{C}(\mathbb{S}^*_0) \geqslant N-\ell+1\right\} \leqslant \int_0^\infty \left(p_{\text{Gauss}} \star p_{\text{Gumbel}}\right)\left(r; \mu_{\text{U}_\ell}, \sigma_{\text{U}_\ell}; \kappa_{\text{L}}, \tau_{\text{L}}\right) \mathrm{d}r, \quad (113)$$

where $p_{\text{Gumbel}}(r;\kappa,\tau)$ *is the Gumbel asymptotic extreme value distribution*

$$p_{\text{Gumbel}}(r;\kappa,\tau) \triangleq \frac{1}{\tau} \cdot \exp\left[-\frac{r-\kappa}{\tau} - \exp\left(-\frac{r-\kappa}{\tau}\right)\right], \quad (114)$$

the integrand $(p_{\text{Gauss}} \star p_{\text{Gumbel}})(r;\mu,\sigma;\kappa,\tau)$ *is the cross-correlation function of the Gaussian and Gumbel pdfs, and all the remaining parameters are defined or detailed below*

$$\mu_{\text{U}_\ell} \triangleq \lambda_\ell + \sqrt{\tfrac{2}{\pi}} \sum_{\substack{j=1 \\ j\neq i}}^{\eta} \sqrt{R_{\ell j}} + (N-\eta)\sigma_y^2 L_{1/2}^{\eta/2-1}\left(-(N-\eta)^2\sigma_y^2\lambda_\ell/\pi\right), \quad (115)$$

$$\sigma^2_{\text{U}_\ell} \triangleq \eta + 4\sigma_y^2\lambda_\ell + \left(1-\tfrac{2}{\pi}\right)\sum_{\substack{j=1 \\ j\neq \ell}}^{\eta} R_{\ell j} + \tfrac{2\lambda_\ell}{\pi}(N-\eta)^2\sigma_y^2 - \mu^2_{\text{U}_\ell}, \quad (116)$$

$$R_{\ell j} \triangleq (\eta\sigma_y^2 + \lambda_\ell + \lambda_j)\sigma_y^2, \quad (117)$$

$$\kappa_{\text{L}} \triangleq P^{-1}_{\text{Gamma}}\left(1 - \tfrac{1}{N-\eta}; m_{\text{L}}, R_{\text{L}}\right), \quad (118)$$

$$\tau_{\text{L}} \triangleq P^{-1}_{\text{Gamma}}\left(1 - \tfrac{1}{(N-\eta)e}; m_{\text{L}}, R_{\text{L}}\right) - \kappa_{\text{L}}, \quad (119)$$

$$R_{\text{L}} \triangleq \mathcal{R}(2; \{m_k\}, \{R_k\}), \quad (120)$$

$$m_{\text{L}} \triangleq \frac{R_{\text{L}}^2}{\mathcal{R}(4; \{m_k\}, \{R_k\}) - R_{\text{L}}^2}, \quad (121)$$

$$\mathcal{R}(n; \{m_k\}, \{R_k\}) \triangleq n! \cdot \sum_{\{q_1,\cdots,q_K\}} \prod_{k=1}^{K} \frac{\mathcal{M}_{q_k}(m_k, R_k)}{q_k!}, \quad (122)$$

$$m_1 \triangleq 1; \quad R_1 \triangleq 8(\eta^2+2\eta)\sigma_y^4, \quad (123)$$

$$m_2 \triangleq \eta/2; \quad R_2 \triangleq \tfrac{4}{\pi}(N-\eta-1)^2\sigma_y^4, \quad (124)$$

$$m_{k>2} \triangleq 1/2; \quad R_{k>2} \triangleq (\eta\sigma_y^2 + \lambda_{k-2})\sigma_y^2, \quad (125)$$

in which $L^\nu_{1/2}(r)$ *is used to represent the generalized Laguerre polynomial of degree* $1/2$*,* $P_{\text{Gamma}}(r;m,R)$ *denotes the Gamma cdf, the summation in equation* (122) *is taken over all possible* K*-tuples satisfying the Diophantine constraint* $q_1 + \cdots + q_K = n$*,* $\mathcal{M}_{q_k}(m_k, R_k)$ *is the moment function defined in equation*

(251) *and* $(\{m_k\}, \{R_k\})$ *are the* K*-tuples of parameters described in equations* (123)-(125).

Before offering a proof for the above theorem, let us explain that the slight abuse of notation committed (for the sake of notational simplicity and without sacrifice of clarity) in the symbols $R_{\ell j}$, R_{L} and R_k is motivated by the fact (as revealed in the proof) that all these quantities are indeed second raw moments of Nakagami-m variates[53].

The different subscripts utilized are respectively meant to clarify the dependence on λ_ℓ and λ_j, the reference to the lower $N-\eta$ rows of $\mathbf{G}^*_{\bar{T}}$, and the indexation with respect to k as detailed in equations (123)-(125).

Proof. Start by invoking Theorem 1, which establishes that the cardinality of a superdisc is given by the number of eigenvalues it encloses. It follows that the cardinality of the null superdisc $\mathbb{S}^*_0$, which is at least $N-\eta$ since it encloses all null eigenvalues of $\mathbf{G}^*_{\bar{T}}$, is augmented by $\eta-\ell+1$ with the probability that the ℓ-th largest eigenvalue is also enclosed by $\mathbb{S}^*_0$. This probability is upper-bounded by the probability that the rightmost point of $\mathbb{S}^*_0$ is not inferior to the leftmost point of the disc $\mathbb{D}_i$ that encloses λ_ℓ, *e.i*,

$$\Pr\left\{\mathcal{C}(\mathbb{S}^*_0) \geqslant N-\ell+1\right\} \leqslant \Pr\left\{\max_{\forall i>\eta} \upsilon_i + \bar{\omega}_i \geqslant \upsilon_\ell - \bar{\omega}_\ell\right\}. \tag{126}$$

The latter probability can be evaluated by invoking the results above. Indeed, from the definition of Gershgorin discs, Lemma 2 and equation (98a) it is seen that $\upsilon_\ell - \bar{\omega}_\ell$ is lower-bounded by

$$\upsilon_\ell - \bar{\omega}_\ell \geqslant \rho_\ell - \sum_{\substack{j=1\\ j\neq i}}^{\eta} |\phi_{\ell j}| - \sum_{j>\eta}^{N} |\theta_{\ell j}|, \quad \ell \leqslant \eta. \tag{127}$$

In the above, ρ_ℓ is Gaussian-distributed (see equation (93), Lemma 2), $\sum_{j>\eta}^{N} |\theta_{\ell j}|$ follows a scaled non-central χ distribution (see equation (109), Corollary 1) and $|\phi_{\ell j}|$ are Nakagami-m variates (see equation (111), Lemma 5).

By force of the Central Limit Theorem, the righthand side of inequality (127) is well approximated by a Gaussian-distributed variate, which due to the independence amongst ρ_ℓ, $|\phi_{\ell j}|$ and $\sum_{j>\eta}^{N} |\theta_{\ell j}|$, has mean μ_{U_ℓ} and variance $\sigma^2_{\mathrm{U}_\ell}$

[53]The same is true for the parameter R_ϵ appearing in Lemma 3.

given by the sum of the means and variances of the corresponding components. Thus,

$$(\upsilon_\ell - \bar{\omega}_\ell) \sim p_{\text{Gauss}}(r; \mu_{\text{U}_\ell}, \sigma_{\text{U}_\ell}), \tag{128}$$

where μ_{U_ℓ} and variance $\sigma^2_{\text{U}_\ell}$ can be shown from equations (93), (109) and (111) to be as described in equations (115) and (116).

Likewise, from the definition of Gershgorin discs, Lemma 3 and equation (98b) we have

$$\upsilon_i + \bar{\omega}_i \leqslant \epsilon + \sum_{j=1}^{\eta} |\theta_{ji}| + \sum_{\substack{j>\eta \\ j\neq i}}^{N} |\psi_{ij}|, \quad i > \eta. \tag{129}$$

In other words, each variate $\upsilon_i + \bar{\omega}_i$, with $i > \eta$, is upper bounded by a sum of Nakagami-m variates whose parameters can be shown from equations (95), (110) and (112) to be as described[54] in equations (123)-(125). In turn, it has been shown in [153] that the sum of Nakagami-m variates is itself well approximated by a Nakagami-m variate, whose parameters can be computed by definition based on the moments of the variates summed utilizing the multinomial theorem, which leads to the expressions provided in equations (120) and (121).

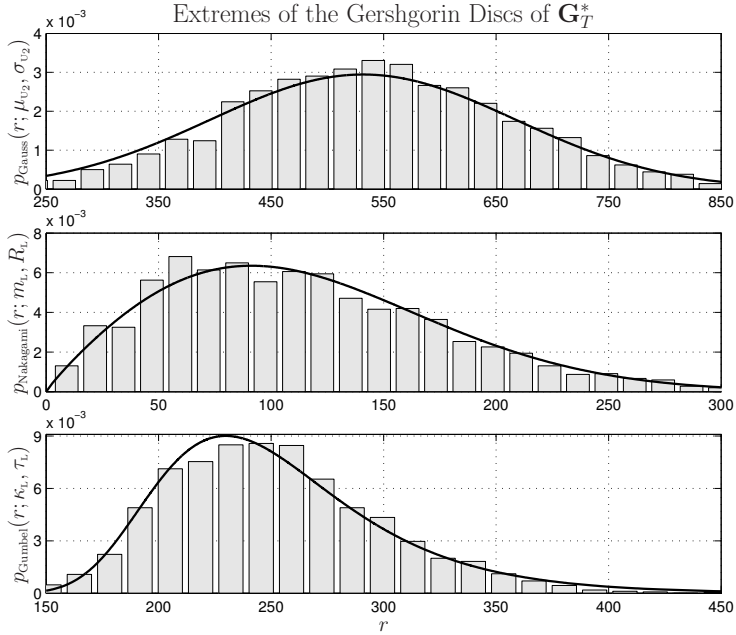
Returning to inequality (126), recall that we seek to characterize the probability that the largest of the $N - \eta$ distinct $(\upsilon_i + \bar{\omega}_i)$'s obtained with a given realization of $\mathbf{Y}$ is larger than $\upsilon_\ell - \bar{\omega}_\ell$. In other words, we are concerned with the extreme value statistics [154] of $N - \eta$ independent Nakagami-m variates. Fortunately, it has been shown [155] that the asymptotic extreme value statistics of Nakagami-m variates are described by the Gumbel distribution defined in equation (114). Thus,

$$\max_{\forall i>\eta} \{\upsilon_i + \bar{\omega}_i\} \sim p_{\text{Gumbel}}(r; \kappa_{\text{L}}, \tau_{\text{L}}), \tag{130}$$

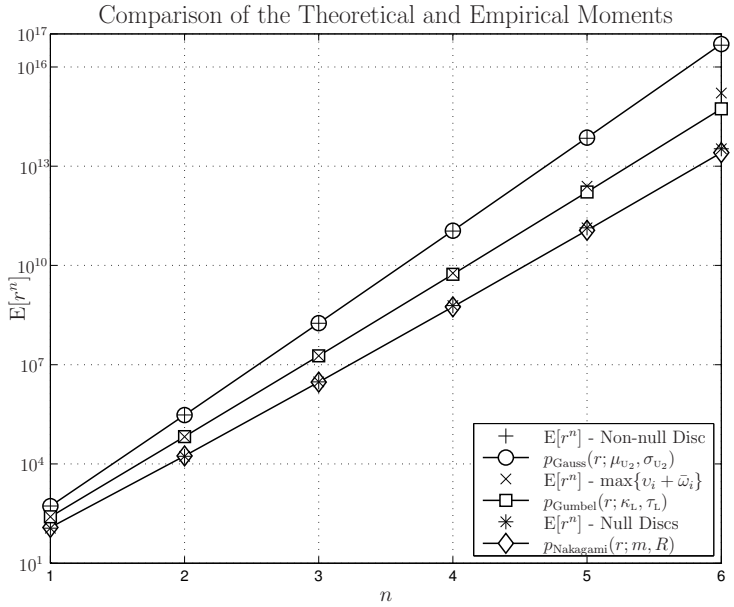
where the parameters κ_{L} and τ_{L}, computed from [155, Lem. 1] and equations (123), (124) and (125), are as given in equations (118) and (119).

It follows that the probability on the righthand side of inequality (126) is in the form $\Pr\{r_{\text{Gumbel}} - r_{\text{Gauss}} \geqslant 0\}$, where the distribution of the variate difference

[54] Recall that ϵ is a bound in probability of ε_i, with bounding probabilities increasing fast with the second-moment ratio $\sqrt{\text{E}[\epsilon^2]/\text{E}[\varepsilon^2]}$, as illustrated in Figure 11.



(a) Comparison between the theoretical models derived in Section 4.2.4 for the extremes of the Gershgorin discs of $\mathbf{G}_T^*$ and the empirical distribution obtained from random data.



(b) Comparison of the theoretical and empirical moments for the extremes of the Gershgorin discs of $\mathbf{G}_T^*$.

Fig 12. Validation of statistical models of Section 4.2. Revised from [12].

is given by the cross-correlation of corresponding pdf's, *i.e*, $(r_{\text{Gumbel}} - r_{\text{Gauss}}) \sim (p_{\text{Gauss}} \star p_{\text{Gumbel}})(r; \mu, \sigma; \kappa, \tau)$, from which equation (113) results immediately, concluding the proof. □

Theorem 2 provides a detailed recipe to evaluate analytically the likelihood and the extent to which a Gramian matrix $\mathbf{G}_{\bar{T}}$ is brought to a *quasi*-diagonal form by the similarity transformation $\mathbf{V}^{\mathrm{T}} \cdot \mathbf{G}_{\bar{T}} \cdot \mathbf{V}$, where $\mathbf{V}$ is the diagonalizer of a matrix $\mathbf{G}$ whose subspace distance to $\mathbf{G}_{\bar{T}}$ is given by $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$. To be precise, equation (113) yields the probability that the null superdisc $\mathbb{S}_0^*$ includes the ℓ-th largest non-zero eigenvalue of $\mathbf{G}_{\bar{T}}^*$. Consequently, if one sets $\ell = \eta$, the probability that the null superdisc includes at least one (the smallest) non-zero eigenvalue results. Notice that such a probability is the likelihood that the eigen-spectrum of $\mathbf{G}_{\bar{T}}$ cannot be obtained by processing (see Section 4.3) only the first η rows of $\mathbf{G}_{\bar{T}}^*$.

Let us now illustrate the accuracy of the analysis carried out in this section.

First, consider Figure 12(a) which illustrates the accuracy of the two models for the statistics of the bounding variates appearing on the righthand side of inequalities (127) and (129), as well as the extreme value distribution of the quantity on the lefthand side of equation (130). Specifically, the upper plot in the figure shows the superposition of a Gaussian pdf with mean and variance as given by equations (115) and (116), respectively, against the empirical distribution of the left limit $v_2 - \bar{\omega}_2$ of the Gershgorin discs associated with the second-largest eigenvalues of various realizations of $\mathbf{G}_{\bar{T}}^*$ constructed using equations (79) and (80) with $N = 25$, the random dislocation matrices given by $\mathbf{Y} = [y_{ij}]$ with $y_{ij} \sim p_{Gauss}(r; 0, \sigma_y = 1)$, and the constant location matrix given by $\mathbf{X} = [x_{ij}]$ with $x_{ij} \triangleq (5(i-1), 5(j-1))$.

Likewise, the middle plot in Figure 12(a) shows a superposition of the Nakagami-m model for the righthand side of inequality (129) for arbitrary $i > \eta$, with R_{L} and m_{L} as in equations (120) and (121), against the empirical distribution obtained from random data. And finally, the lower plot compares the empirical distribution of the largest Gershgorin right limit $v_{i>2} - \bar{\omega}_{i>2}$ observed in each realization, against the corresponding Gumbel extreme value distribution of equation (114) with parameters as in equations (118) and (119).

To further reinforce the accuracy of the statistical models as indicated by Figure 12(a), a comparison of the theoretical and empirical moments of

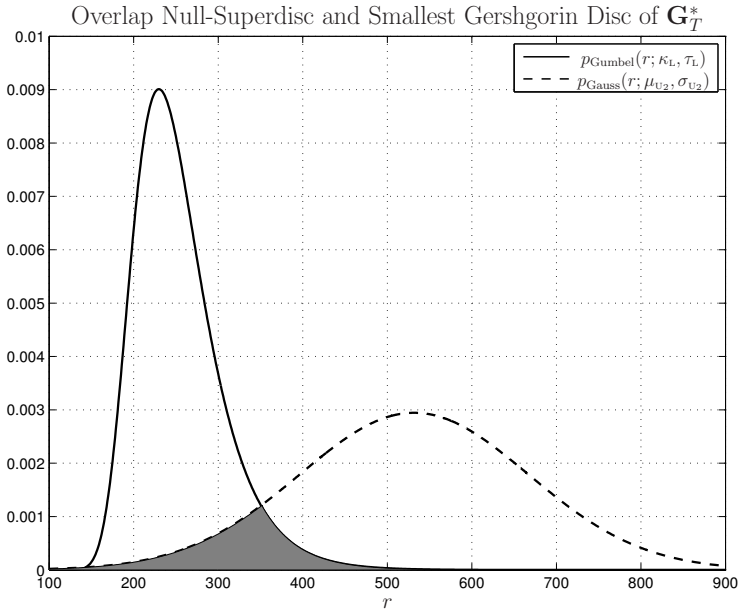


Fig 13. Representation of the probability of overlap between the null-superdisc $\mathbb{S}_0^*$ and the smallest non-null Gershgorin discs of $\mathbf{G}_{\bar{T}}^{*}$ as a function of $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ for the grid scenario described in Section 4.4. Revised from [16].

corresponding random variables are also shown in Figure 12(b). Overall Figure 12 demonstrates that the two critical Gershgorin limits that determine the likelihood that the null superdisc includes the non-zero ℓ-th largest eigenvalue – namely $\upsilon_\ell - \bar{\omega}_\ell$ and $\max_{\forall i > \eta} \{\upsilon_i + \bar{\omega}_i\}$ – are accurately described by the models indicated in equations (128) and (130), respectively.

An illustration of integral in equation (113) is given in Figure 13, while Figure 14 illustrates the computation of the bound accuracy of the bound given by equation (113) put forward in Theorem 2 by comparing the analytically evaluated probability of overlap between the ℓ-th Gershgorin disc and the null superdisc $\mathbb{S}_0^*$. The figure confirms that the various bounding procedures adopted in the analysis thus far, without which the relatively simple result obtained would hardly be possible, do not undermine its accuracy.

While it is understandable that readers may wish to draw a parallel between the quantity $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ – which albeit the relevant one, is indeed abstract – and

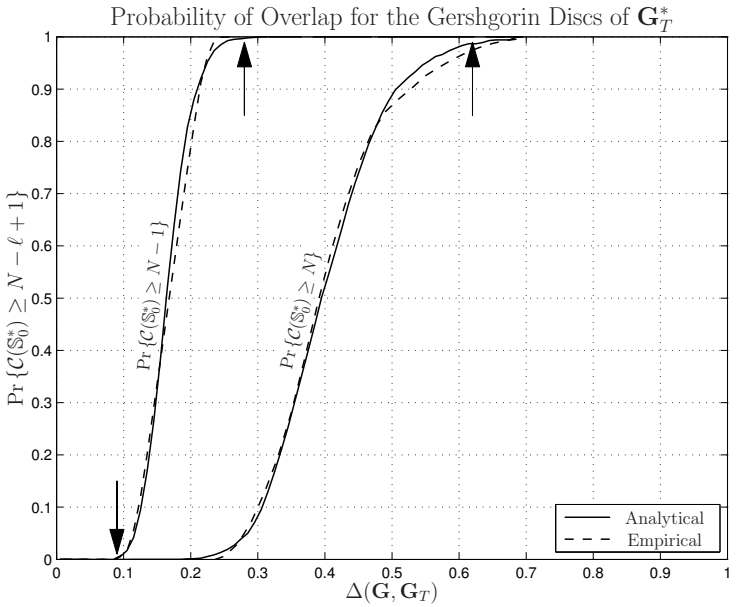


Fig 14. Probability of overlap between the null-superdisc of $\mathbf{G}_{\bar{T}}^*$ and its non-null Gershgorin discs, as a function of $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ for the grid scenario described in Section 4.4. Revised from [17].

tangible physical parameters, such a relationship is non-trivial in face of the fact that $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ actually depends on multiple factors, including the progress of targets between samples, but also the size of the matrix, the noise power and the actual location of targets ($\mathbf{X}$) at the sampling moment.

This can be better understood from Figure 15, where $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ is plotted as a function of the dislocation variance σ_y (which captures the effects of velocity, noise, sampling period, etc).[55]

Finally, Theorem 2, seen through the prism of MDS-based multitarget tracking, tells us that the information on the eigen-spectrum of dynamic MDS kernels lays, with high probability (also quantified by the Theorem), in the first η rows of $\mathbf{G}_{\bar{T}}^*$. In other words, Theorem 2 suggests that the eigen-decomposition of $\mathbf{G}_{\bar{T}}^*$ can be done efficiently and accurately at a fraction of the cost, by focusing on a selection of elements. The latter is the subject of the next section.

[55] Curves obtained for a "grid" scenario described in Section 4.4. Different realizations of $\mathbf{X}$ would obviously yield slightly different relationships between $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ and σ_y.

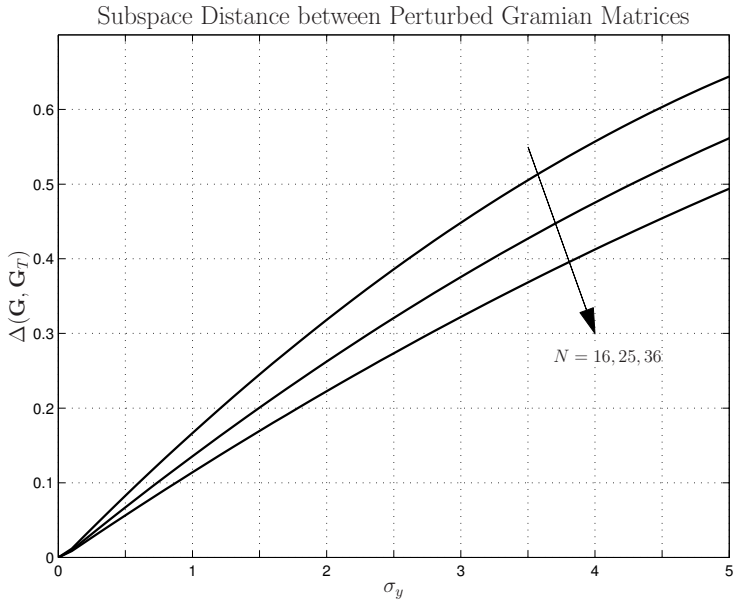


Fig 15. Subspace distance $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ as a function of σ_y for the grid scenario described in Section 4.4.

4.3 Jacobi MDS-based multitarget tracking

Let us put the results presented thus far into the context of MDS-based multitarget tracking. Specifically, our objective is to show how the results above can be used to optimize the eigen-decomposition of dynamic Euclidean kernels arising in multitarget tracking schemes as described in Section 4.1.

4.3.1 Jacobi and Gershgorin over Gramian matrices

First, notice that one implication of the Gershgorin Theorem (Theorem 1) is that if a row of $\mathbf{G}^*_{\bar{T}}$ has an associated disc disjoint to all other discs[56], then that row determines the "*location*" (within a margin of error) of one of the eigenvalues of $\mathbf{G}^*_{\bar{T}}$. In other words, the Gershgorin Theorem establishes a direct, one-to-one relationship between the given row and the corresponding eigenvalue of $\mathbf{G}^*_{\bar{T}}$. Therefore, what is required in order to exploit this feature of the Gershgorin

[56] Accordingly, if a set of rows is associated to a superdisc disjoint to all other discs.

bound is an algorithmic tool to, loosely speaking, process the entries of the given row of $\mathbf{G}^*_{\bar{T}}$ into the corresponding eigenvalue.

Such tool is the Jacobi-Givens rotation. Indeed, in the context of MDS-based tracking, the eigenvector matrix $\mathbf{V}_{\bar{T}}$ of $\mathbf{G}^*_{\bar{T}}$ is given by

$$\mathbf{V}_{\bar{T}} = \left(\prod_{s=1}^{N_{\mathrm{S}}} \prod_{(i_s,j_s)} \mathcal{R}(i_s, j_s, \vartheta) \right) \cdot \mathbf{V}, \tag{131}$$

where N_{S} denotes the number of times the matrix is "swept" by plane rotations[57] and $\mathcal{R}(i,j,\vartheta)$ is the Jacobi-Givens rotation matrix[58] associated to a quartet $\{k_{ii}, k_{ij}, k_{ji}, k_{jj}\}$ of a generic symmetric matrix $\mathbf{K} \in \mathbb{R}^{N\times N}$.

Now, define the quantities

$$\varphi_{ij}(\mathbf{K}) \triangleq k_{ii}^2 + k_{jj}^2, \tag{132}$$

$$\varpi_{ij}(\mathbf{K}) \triangleq 2\Bigg(\sum_{\substack{\ell=1 \\ \ell\neq i}}^{N} k_{i\ell}^2 + \sum_{\substack{\ell=1 \\ \ell\neq j}}^{N} k_{j\ell}^2 \Bigg). \tag{133}$$

Then, the following applies[59] [156, pp. 178]

$$\varphi_{ij}\left(\mathcal{R}(i,j,\vartheta)^{\mathrm{T}} \cdot \mathbf{K} \cdot \mathcal{R}(i,j,\vartheta)\right) = \varphi_{ij}\left(\mathbf{K}\right) + 2k_{ij}^2, \tag{135}$$

$$\varpi_{ij}\left(\mathcal{R}(i,j,\vartheta)^{\mathrm{T}} \cdot \mathbf{K} \cdot \mathcal{R}(i,j,\vartheta)\right) = \varpi_{ij}\left(\mathbf{K}\right) - 2k_{ij}^2. \tag{136}$$

Clearly φ_{ij} and ϖ_{ij} relate to the centers and the radii of the i-th and j-the Gershgorin discs $\mathbb{D}_i$ and $\mathbb{D}_j$, respectively. In particular, equations (135) and (136) indicate that each Jacobi-Givens rotation acts upon $\mathbb{D}_i$ and $\mathbb{D}_j$, dislocating their centers and shrinking their radii.

In general one cannot infer any further as to how the centers of $\mathbb{D}_i$ and $\mathbb{D}_j$ are dislocated, individually. In the case where the matrix $\mathbf{K}$ under consideration is the positive semidefinite $\mathbf{G}^*_{\bar{T}}$, and the indexes i and j are such that $i \leqslant \eta$ and

[57] In other words, $N_{\mathrm{S}} = 1$ indicates $N(N-1)/2$ plane rotations.

[58] A concise definition of the elements of $\mathcal{R}(i,j,\vartheta)$, given the quartet $\{k_{ii}, k_{ij}, k_{ji}, k_{jj}\}$, is [143]

$$[\mathcal{R}(i,j,\vartheta)]_{\ell q} \triangleq (\ell \perp q) \cdot \cos[(\ell\perp i + \ell\perp j)\cdot\vartheta] + (\ell\perp i)\cdot\sin[(q\perp j)\cdot\vartheta] - (\ell\perp j)\cdot\sin[(q\perp i)\cdot\vartheta], \tag{134}$$

where $\vartheta \triangleq \frac{1}{2}\cdot\mathrm{atan}\left(\frac{k_{ij}+k_{ji}}{k_{ii}-k_{jj}}\right)$ with $\mathrm{atan}(\cdot)$ defined between $(-\pi/4, \pi/4]$, and where $\perp$ denotes Boolean equality such that $\ell\perp q \triangleq 1$ if $\ell = q$ and 0 otherwise.

[59] The result shown here is slightly stronger than that in [156, pp. 178] and follows from the latter combined with the symmetry of Gramian matrices and the fact that the Jacobi rotation performed by $\mathcal{R}(i,j,\vartheta)$ annihilates the element k_{ij}^2.

$j > \eta$, however, we can infer – by force of Theorem 2, the absolute and step-wise convergence of the Jacobi algorithm [157, 158] and equations (132) and (135) – that with a high probability every Jacobi-Givens rotation yields

$$\tilde{k}_{ii}^2 = k_{ii}^2 + 2k_{ij}^2 + c, \tag{137}$$

$$\tilde{k}_{jj}^2 = k_{jj}^2 - c, \tag{138}$$

where $\tilde{k}_{ii}$ and $\tilde{k}_{jj}$ represent here the i-th and j-th diagonal element of $\mathcal{R}(i,j,\vartheta)^{\mathrm{T}} \cdot \mathbf{K} \cdot \mathcal{R}(i,j,\vartheta)$, and c is an unknown *positive* number.

In other words, in the case under consideration, the impact of Jacobi-Givens rotations on the centers of $\mathbb{D}_i$ and $\mathbb{D}_j$ is understood, separately, as a consequence of the facts that, under the conditions outlined above, k_{ii} and k_{jj} are positive and k_{jj} is brought closer to 0 with every Jacobi-Givens rotation, since Theorem 2 establishes that with a high probability the bottom $N - \eta$ rows of $\mathbf{G}_{\bar{T}}^*$ is associated with null eigenvalues only.

4.3.2 Statistical optimization of Jacobi algorithm

It is known that quadratic convergency (on the number of plane rotations) is ensured if every Jacobi-Givens rotation is optimized by selecting the quartet with the largest off-diagonal element [156]. Unfortunately, however, this "conventional" optimal sweeping strategy is somewhat computationally demanding since it requires that *all* the off-diagonal elements of the matrix to be decomposed be compared before each plane rotation.

This drawback of the conventional Jacobi algorithm partially motivates the effort to devise other alternatives (see *e.g.* [159, 160]), which typically offer lower complexity at the expense of limitations in terms friendliness to parallel/distributed processing, numerical stability, etc.

Fortunately, once again results from Section 4.2 can be invoked to circumvent this problem. Indeed, recall that:

- The distribution of the diagonal entries of the first η rows of $\mathbf{G}_{\bar{T}}^*$ is described by equations (93) and (92) of Lemma 2
- The distribution of the diagonal entries of the bottom $N - \eta$ rows of $\mathbf{G}_{\bar{T}}^*$ is described by equations (95) and (96) of Lemma 3
- The distributions of the off-diagonal entries of the bottom $N - \eta$ rows of $\mathbf{G}_{\bar{T}}^*$

are described by equations (248) and (249) of Lemma 4
– The distributions of the off-diagonal entries of the first η rows of $\mathbf{G}^*_{\bar{T}}$ are described by equations (111) and (112) of Lemma 5.

It is important to highlight that in possession of this information, we are able to determine *a priori*, in a statistical sense, which elements of $\mathbf{G}^*_{\bar{T}}$ are likely to be largest and therefore should be chosen first to compose the quartets over which $\mathcal{R}(i_s, j_s, \vartheta)$ are constructed. Namely

1. The elements of the first η-by-η minor
2. The remaining elements of the first η rows.

Notice also that in light of Theorem 2, Jacobi-Givens rotations over this smaller portion of $\mathbf{G}^*_{\bar{T}}$ ensures (with high probability) that all the non-zero eigenvalues and eigenvectors of $\mathbf{G}^*_{\bar{T}}$ are found, such that tracking accuracy is not in anyway compromised.

To conclude this subsection let us point out that the Jacobi algorithm also offers several advantageous features of interest, such as: accuracy [161, 162]; numerical stability [156]; friendliness to parallel and distributed-processing [163, 164]; and generalizability to joint-diagonalization applications [157, 165].

4.4 Performance evaluation and comparisons

In order to evaluate the performance of the Jacobi-based technique described in Section 4.3, let us consider the number of plane rotations N_{S} required by each sweeping strategy in order to bring the sum of the off-diagonal terms of $\mathbf{G}^*_{\bar{T}}$ to "zero", that is, a reasonable machine epsilon [166] which we set to 10^{-8}.

Specifically, let us consider the following sweeping strategies for the Jacobi algorithm:

1. *Cyclic*: naïve strategy in which sweeps are performed sequentially, row by row
2. *Optimum*: at each rotation, search and annihilate the element with the largest magnitude
3. *Largest Mean*: apply rotations designed to annihilate the elements of $\mathbf{G}^*_{\bar{T}}$ characterized by largest moments
4. *Truncated LM*: constrain the largest mean (LM) strategy to the first η rows of $\mathbf{G}^*_{\bar{T}}$.

For the sake of completeness, the performances of these sweeping strategies will be compared to the Symmetric QR algorithm, which is know to be equivalent to a Jacobi algorithm with $N_{\mathrm{S}} \approx 2$ complete sweeps [143, pp. 431]. These eigen-decomposition techniques are applied to MDS kernels obtained as *per* equations (79) and (80) with $N = 25$. The constant location matrix $\mathbf{X} = [x_{ij}]$ with $x_{ij} \triangleq (5(i-1), 5(j-1))$ – which yields a grid topology – and multiple realizations of the random dislocation matrices are generated with $\mathbf{Y} = [y_{ij}]$ with $y_{ij} \sim p_{Gauss}(r; 0, \sigma_y)$ and $(0 < \sigma_y < 25)$.

The results are summarized in Table 1. As expected, the results indicate that the penalty of replacing the search for adequate (large) off-diagonal entries before each rotation with a naïve cyclic sweeping is that a larger number of sweeps is required, such that the cyclic strategy is more costly than the QR algorithm.

In comparison, the optimum sweeping requires the least number of sweeps amongst all alternatives. It must be taken into consideration, however, that such a strategy relies on an exhaustive search for the largest off-diagonal element in the matrix before each plane rotation is applied, and therefore is highly costly. The statistically optimum strategy based on performing rotations to annihilate the entries of $\mathbf{G}_{\bar{T}}^{*}$ with the largest mean, which is made possible by exploiting the statistical characterization of the entries in $\mathbf{G}_{\bar{T}}^{*}$ given in Lemma 5, allows for eliminating the need for exhaustive searches while maintaining the number of required rotations below 2, thus yielding an algorithm less complex than the symmetric QR.

Table 1. Number of plane rotations for the Jacobi algorithm.

$\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$	Number of Sweeps - N_{S} - to decompose $\mathbf{G}_{\bar{T}}^{*}$				
	S. QR[a]	*Cyclic*	*Optimum*	*Largest Mean*	*Truncated LM*
0.1	2	2.65	0.24	1.12	0.24
0.2	2	2.65	0.26	1.17	0.28
0.3	2	2.58	0.28	1.20	0.31
0.4	2	2.59	0.30	1.21	0.31
0.5	2	2.62	0.30	1.60	0.31

[a] For the symmetric QR algorithm $N_{\mathrm{S}} \approx 2$ [143, pp. 431].

Finally, invoking Theorem 2 the probability that the Gershgorin discs

associated to the η-dimensional dominant eigenspace of $\mathbf{G}^*_{\bar{T}}$ – which suffice for MDS-based tracking application – are disjoint from the null superdisc of the matrix can be inferred. In other words, Theorem 2 provides an estimate of the confidence that Jacobi algorithm converges to the η dominant eigenspace[60] of the matrix if rotations are confined to the first η rows of $\mathbf{G}^*_{\bar{T}}$, which is the principle behind the truncated LM strategy.

It is seen from Table 1 that this new strategy only requires a fraction of the original complexity, in fact achieving a complexity comparable to that of the optimum strategy. But since the entries upon which the rotation matrix $\mathcal{R}^*(i_s, j_s, \vartheta)$ is to be applied is chosen *a priori*, no exhaustive search is required.

The trend seen in Table 1 is further corroborated by the results shown in Figure 16 and 17 where the number of plane rotations required by the optimum and the truncated LM strategy are compared for the entire span of values of $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$. For comparison purposes, the figure also shows the performance of the optimum strategy applied to $\mathbf{G}_{\bar{T}}$, from which it becomes evident that the similarity transformation defined in equation (80) not only is never harmful.

Indeed, comparing Figure 14 and 16 it is found that $\Pr\{\mathcal{C}(\mathbb{S}^*_0) \geqslant N - \ell + 1\}$ and the number of plane rotations required by the truncated LM strategy undergo the same transients at the same values of $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$, which explains why the proposed sweeping strategy fully captures the relation existing between subsequent Gram matrices.

Notice that Theorem 2 describes the Gershgorin spectral bounds of $\mathbf{G}^*_{\bar{T}}$, thus accounting only for the similarity transformation defined in equation (80), not accounting for the changes on the statistics of the entries of the matrix being decomposed after each plane rotation. Nevertheless, Figures 14 and 16 indicate that the number of plane rotations required to eigen-decompose $\mathbf{G}^*_{\bar{T}}$ is properly captured by the analysis. In particular, for the scenario considered, it has been verified that the truncated criterion yields the exact dominant eigenspace of $\mathbf{G}_{\bar{T}}$ up to $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}}) \approx 0.75$. In addition, by checking the diagonal terms of $\mathbf{G}^*_{\bar{T}}$ it is possible to retrieve the eigenspace of $\mathbf{G}^*_{\bar{T}}$ even for $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}}) > 0.75$ at the price of only a slight increase in complexity. The results shown in Figure 16, 17 and 18 as well as in Table 1 use this modification.

[60] Although the truncated LM strategy only recovers the η dominant eigenspace of $\mathbf{G}^*_{\bar{T}}$, the null-eigenspace of the matrix can be computed at the price of an additional but limited number of plane rotations by simply making the LM strategy to be followed by cyclic sweeping strategy.

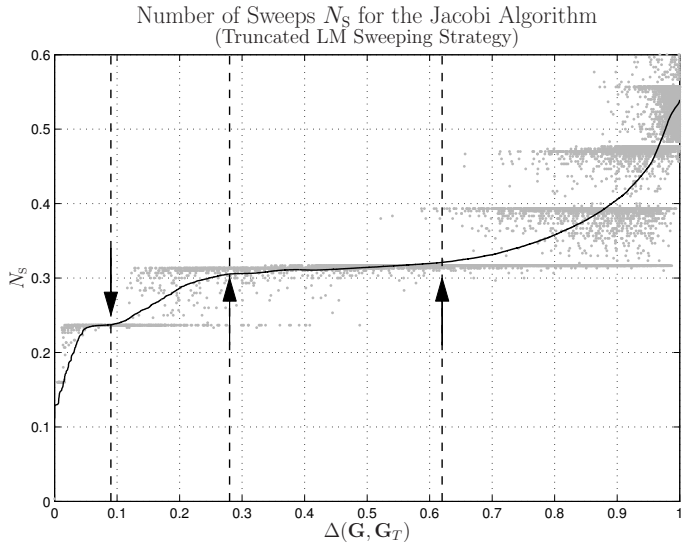


Fig 16. Complexity required by the truncated LM sweeping strategy to eigen-decompose $\mathbf{G}^*_{\bar{T}}$. Results are shown as a function of $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ for the grid scenario described in Section 4.4. Revised form [12].

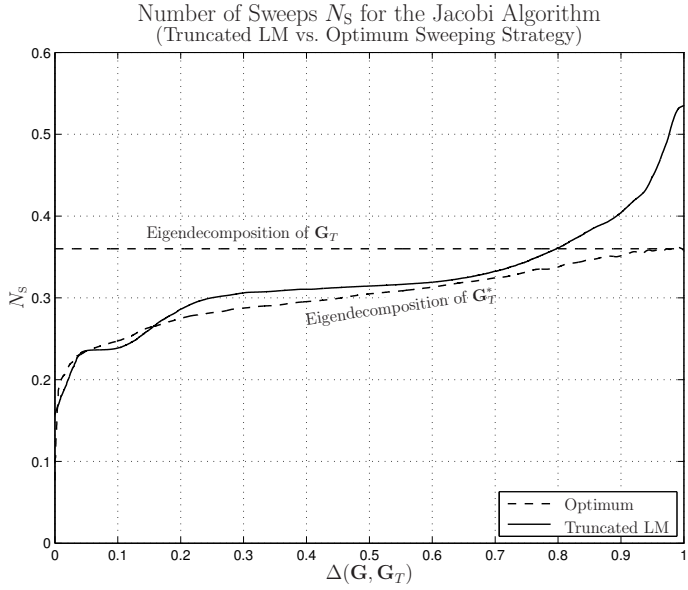


Fig 17. Complexity required by the truncated LM and optimum sweeping strategies to eigen-decompose $\mathbf{G}^*_{\bar{T}}$, as a function of $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$ for the grid scenario described in Section 4.4. Revised form [12].

Next, the complexity – measured in terms of the number of sweeps normalized by the subspace distance ($N_{\mathrm{S}}/\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$) – of the Truncated LM algorithm as a function of the size of the matrix (squared-rooted) is shown in Figure 18. It is found that independently on the standard dislocation σ_y affecting the entries of $\mathbf{Y}$ in equation (79), the truncated LM algorithm achieves a remarkably low complexity of order $\mathcal{O}(\sqrt{N})$.

Finally, the performance of the C-MDS solution is compared against the full-set subspace formulation of the MDS problem proposed in [43]. The scenario chosen for this comparison consists of 4 anchor nodes placed at edges of a 10×10 square in a $\eta = 2$ dimensional space and 10 targets having trajectories generated according to first order time homogeneous Markov processes restricted to the 20×20 space centered on the area defined by the anchor's positions. A typical target trajectory is shown in Figure 19. Full connectivity (*i.e.*, full EDM) is assumed amongst all the nodes in the scenario where the anchor-to-target and the target-to-target measurements are perturbed by $\sim p_{Gauss}(r; 0, \sigma_d)$.

The performance metric used in this comparison is the RMSE defined in equation (74) and computed over the location estimates $\hat{\mathbf{X}}$ and averaged over 100 realization of 1000 steps long trajectories. The results are shown in Figure 20 and reveal that for any level of the perturbation σ_d over distance estimates, the C-MDS algorithm is superior to the alternative subspace method.

Moreover, because of the truncated LM strategy utilized in the eigen-decomposition of $\mathbf{G}^*_{\bar{T}}$, the proposed algorithm is able to exploit the subspace distance between subsequent Euclidean kernel to speed-up the computation of the C-MDS algorithm.

Additionally, since the C-MDS algorithm has limited complexity, the computational power saved by utilizing this technique can be employed to improve the robustness of the system against corrupted data by pre-filtering the entries of $\mathbf{D}_{\bar{T}}$. Amongst several possible choices, a low-complexity non-parametric candidate is the Wavelet-based filter used in [14, 167]. Indeed, by decomposing the time series corresponding to the TOA observations through a discrete wavelet transform (DWT) [168] it is possible to use the output of the *scaling* funtion to estimate the target dynamic and the wavelet coefficients to infer on the noise level σ_d.

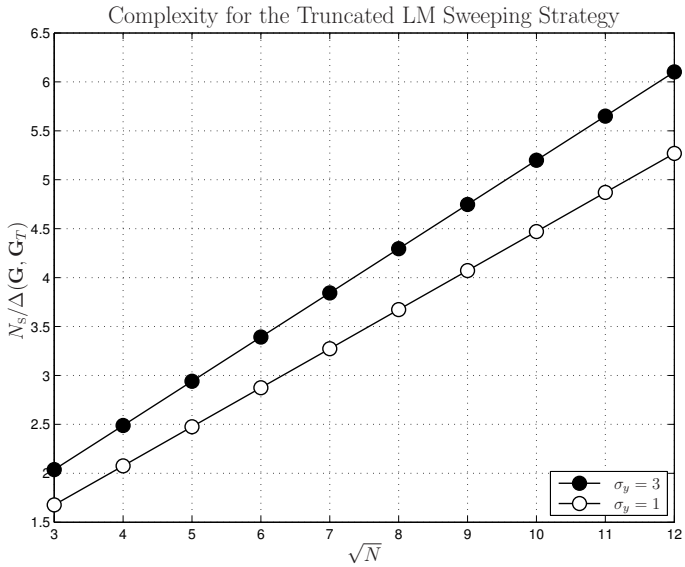


Fig 18. Complexity of the truncated LM sweeping strategy applied to eigen-decompose $\mathbf{G}_T^*$. Results are given for the grid scenario described in Section 4.4 with the dislocation matrix $\mathbf{Y}$ constructed according to two different values of σ_y. Revised form [12].

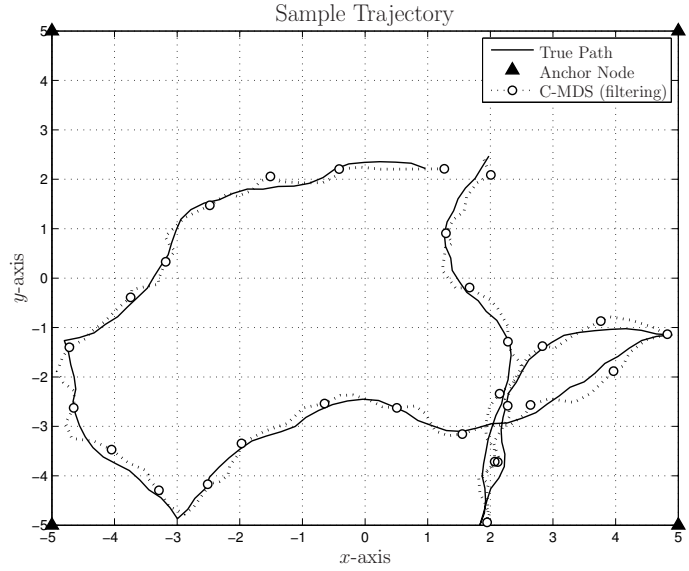


Fig 19. Typical target trajectory and C-MDS estimate with the anchor-to-target measurements perturbed by $\sim p_{\textsf{Gauss}}(r; 0, .5)$. Revised form [12].

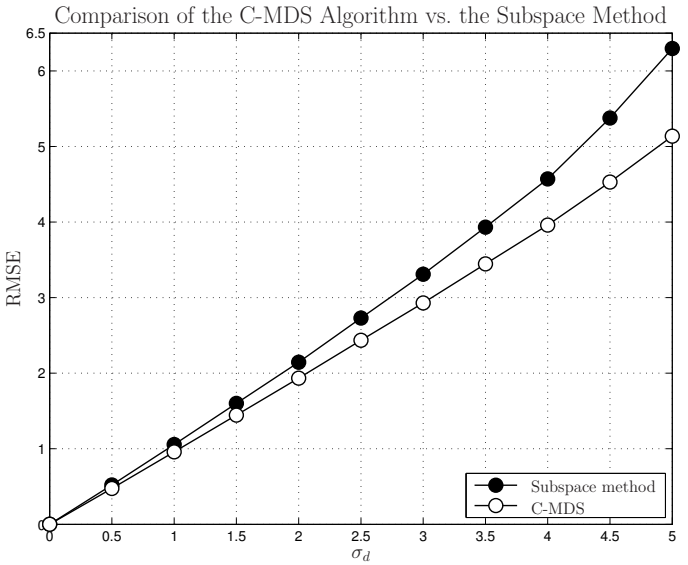


Fig 20. Comparison of the C-MDS algorithm versus the full-set subspace method proposed in [43]. The results are as a function of the noise (σ_d) affecting the range measurements for the tracking scenario described in Section 4.4. Revised form [12].

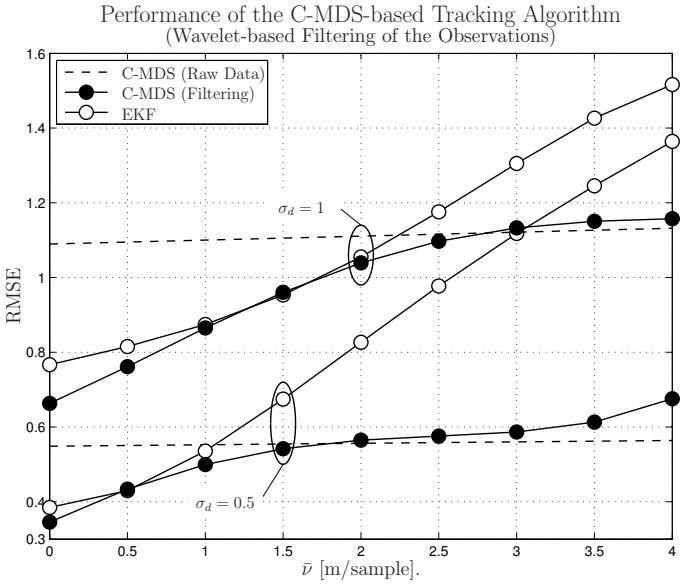


Fig 21. Performance with and without the wavelet-based filtering proposed in [14]. Results as a function of $\bar{\nu}$ for a scenario of 4 anchor nodes placed at the edges of a 10×10 square in a $\eta = 2$-dimensional space. Revised form [12].

The performance achieved by the C-MDS-based tracking algorithm used jointly with the filter described in [14] is compared against an EKF alternative in Figure 21 as a function of the average target dynamic $\bar{\nu}$.

It is important to remember that, unlike the EKF, the filtered C-MDS based tracking algorithm does not rely on any model nor on more than the last few estimates and current observations in order to update the estimates of the location of tracked target.

4.5 Summary and discussions

This chapter proposed a comprehensive algorithm to handle the problem of simultaneously tracking a large number of targets with no *a priori* mobility models, which is accomplished by modifying the Jacobi algorithm used to eigen-decompose the subsequent Euclidean kernels in the MDS algorithm.

Specifically, it was first statistically characterized the spectrum of random $N \times N$ Gramian matrices of the form of $\mathbf{G}^*_{\bar{T}}$ as a function of the subspace distance $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$, where $\mathbf{G}$ and $\mathbf{G}_{\bar{T}}$ are themselves Gramian matrices. To do so, an extreme-value and asymptotic take on the theory of Gershgorin spectrum bounds was employed to characterize the statistical structure of $\mathbf{G}^*_{\bar{T}}$. The results revealed that even for relatively large $\Delta(\mathbf{G}, \mathbf{G}_{\bar{T}})$, the matrix $\mathbf{G}^*_{\bar{T}}$ can, with a high probability (quantified analytically), be eigen-decomposed by employing a statistically optimized number of plane rotations $\mathcal{R}^*(i, j, \vartheta)$. Moreover, we showed that the Jacobi algorithm used together with the proposed truncated LM sweeping strategy achieves the convergence rate of the optimal strategy applied to $\mathbf{G}^*_{\bar{T}}$ but without requiring the search for the optimal element to be annihilated.

All in all, the proposed MDS-based solution combined with the fact that the Euclidean double-centered kernels of the classic MDS method are asymptotically Gramian yields a highly flexible and efficient (fast) MDS-based multitarget tracking algorithm which achieves a remarkably low complexity of order $\mathcal{O}(\sqrt{N})$ and which is immune to the non-stationarity of the target's dynamics.

5 The active MTT problem: incomplete and heterogeneous scenarios

In the following the algebraic framework used in Chapter 4 is extended to cope with scenarios characterized by incomplete EDM and heterogeneous information. Specifically, the problem of randomly sampled EDM matrices is dealt with by incorporating the Nyström approximation of the kernel matrix associated to $\mathbf{D}_{\bar{T}}$, resulting in an efficient and fully algebraic Euclidean Distance Matrix completion technique. It is also shown that, under the assumption that the network is percolated (as required by Graph-based Completion), the same technique can be used in conjunction with low-complexity optimization algorithms as an efficient initialization method, allowing to standard low-complex optimization solutions to perform, similarly to SDP solutions, close to the CRLB.

Heterogeneous scenarios, in which both range and angle informations are assumed to be known, are handled by solving the C-MDS algorithm over the edges of the graph associated to the network of nodes. In so doing both distance and angle information can be processed algebraically (without iteration) and simultaneously. It is shown how by approximating the kernel associated to the graph using the Nyström method it is possible to limit both the complexity as well as the amount of information required by the algorithm, making the approach suitable to real LT applications. Moreover, it is also illustrated how to apply the aforementioned solution to scenarios in which only a subset of the mere range information between the agents is available. The results demonstrate the superiority of the proposed solution compared to the standard C-MDS and the SMACOF algorithms under both LOS and NLOS conditions.

5.1 Incomplete scenarios

One limitation of the C-MDS algorithm discussed in Chapter 4 is that it requires complete knowledge of the Euclidean distance matrix $\mathbf{D}_{\bar{T}}$. However, the fact that the single entries of $\mathbf{K}_{\bar{T}}$ depends on all the entries of $\mathbf{D}_{\bar{T}}$ suggests that the EDM matrix contains redundant information that could be used to cope with the completion problem mentioned above, although the problem of choosing

which elements of $\mathbf{D}_{\bar{T}}$ to use to compute an approximate $\check{\mathbf{K}}$ still remains.

Fortunately this problem can be solved relying on an approximation of the eigen-spectrum of $\mathbf{K}$ computed through the Nyström method. In the following the basics of the Nyström method are presented in Section 5.1.1, while Section 5.1.2 deals on how to use the algorithm in conjunction with the C-MDS framework in LT scenarios to solve the erasure problem in $\mathbf{D}_{\bar{T}}$.

It is also shown that, due to its intrinsically low computational complexity, the proposed solution can be used as an efficient initialization for low-complexity optimization algorithms.

5.1.1 Nyström approximation

The applicability of the C-MDS approach discussed in Chapter 4, can be severely limited by the fact that, in principle, to recover the eigen-pair associated to the MDS kernels $\mathbf{K}$, each single entry of the Euclidean distance matrices $\mathbf{D}_{\bar{T}}$ must be known.

Fortunately, however, there are methods that retrieve approximations of the eigen-decomposition of $\mathbf{K}$ from a subset of its entries only. One such technique is the so called Nyström method [148], which, by applying a quadrature rule evaluated at the set of equally spaced points $\{\varsigma_1, \varsigma_2, \ldots, \varsigma_n\} \in [a, b]$, finds an approximation to the eigen-function problems

$$\int_a^b W(x, y)\, \phi(y)\, dy = \lambda \phi(x), \tag{139}$$

as

$$\frac{(b-a)}{n} \sum_{j=1}^{n} W(x, \varsigma_j) = \lambda \hat{\phi}(x), \tag{140}$$

where $\hat{\phi}(x)$ is an estimate for $\phi(x)$.

Assume $x = \varepsilon_i$, then the system of equations above can be written as

$$\frac{(b-a)}{n} \sum_{j=1}^{n} W(\varsigma_i, \varsigma_j) = \lambda \hat{\phi}(x), \quad \forall i \in \{1, \ldots, n\}. \tag{141}$$

Let $[a, b] = [0, 1]$ in equation (141), then the aforementioned system can be expressed in matrix form as $\mathbf{A}\hat{\mathbf{\Phi}} = n\hat{\mathbf{\Phi}}\mathbf{\Lambda}$,, where $[\mathbf{A}]_{ij} = W(\varepsilon_i, \varepsilon_j)$, while $\mathbf{\Phi} = [\phi_1, \ldots, \phi_n]$ and $\mathbf{\Lambda}$ are respectively the eigenvectors and eigenvalues of $\mathbf{A}$.

It follows that the i-th eigenvector at the arbitrary point x, namely $\phi_i(x)$, can be approximated by $\hat{\phi}_i(x)$ relying only on a subset of points, that is

$$\hat{\phi}_i(x) = \frac{1}{n\lambda_i} \sum_{j=1}^{n} W(x, \varsigma_j)\, \hat{\phi}_i(\varsigma_j). \tag{142}$$

5.1.2 Initialization for non-convex localization problems

As mentioned above, the incompleteness problem in the C-MDS algorithm can be overcome using the *Nyström approximation*[61] [146–148], whose application to the MDS problem can be summarized as follows. First, consider the Nyström kernel given by [148]

$$\breve{\mathbf{K}} \approx \left[\begin{array}{c:c} [\mathbf{K}]_{1:\eta,1:\eta} & [\mathbf{K}]_{1:\eta,\eta+1:N} \\ \hdashline [\mathbf{K}]^{\mathrm{T}}_{1:\eta,\eta+1:N} & [\mathbf{K}]^{\mathrm{T}}_{1:\eta,\eta+1:N} \cdot [\mathbf{K}]^{-1}_{1:\eta,1:\eta} \cdot [\mathbf{K}]_{1:\eta,\eta+1:N} \end{array}\right], \tag{143}$$

in which $[\mathbf{K}]_{1:\eta,1:\eta}$ and $[\mathbf{K}]_{1:\eta,\eta+1:N}$ denote the upper-left $\eta - by - \eta$, and the upper-right $\eta - by - (N-\eta)$ minors of the double center kernel matrix $\mathbf{K}$ defined in equation (18).

Recall also that [169]

$$[\mathbf{K}]_{1:\eta,1:\eta} = -\frac{1}{2} \cdot \left([\mathbf{D}_{\bar{T}}]_{1:\eta,1:\eta} + \mathbf{C}_1 \otimes \mathbf{1}_\eta \mathbf{1}_\eta^{\mathrm{T}} - \mathbf{C}_2 \otimes \mathbf{1}_\eta^{\mathrm{T}} - \mathbf{C}_3 \otimes \mathbf{1}_\eta \right), \tag{144}$$

$$\begin{aligned} [\mathbf{K}]_{1:\eta,\eta+1:N} = -\frac{1}{2} \cdot \big(& [\mathbf{D}_{\bar{T}}]_{1:\eta,\eta+1:N} + \mathbf{C}_1 \otimes \mathbf{1}_\eta \cdot \mathbf{1}^{\mathrm{T}}_{N-\eta} \\ & - \mathbf{C}_2 \otimes \mathbf{1}^{\mathrm{T}}_{N-\eta} - \mathbf{C}_4 \otimes \mathbf{1}_\eta \big), \end{aligned} \tag{145}$$

where $\otimes$ denotes the Kronecker product and

$$\mathbf{C}_1 = \tfrac{1}{\eta^2} \cdot \left[\mathbf{1}^{\mathrm{T}}_\eta \cdot [\mathbf{D}_{\bar{T}}]_{1:\eta,1:\eta} \cdot \mathbf{1}_\eta \right], \tag{146}$$

$$\mathbf{C}_2 = \tfrac{1}{\eta} \cdot \left[[\mathbf{D}_{\bar{T}}]_{1:\eta,1:\eta} \cdot \mathbf{1}_\eta \right], \tag{147}$$

$$\mathbf{C}_3 = \tfrac{1}{\eta} \cdot \left[\mathbf{1}^{\mathrm{T}}_\eta \cdot [\mathbf{D}_{\bar{T}}]_{1:\eta,1:\eta} \right], \tag{148}$$

$$\mathbf{C}_4 = \tfrac{1}{\eta} \cdot \left[\mathbf{1}^{\mathrm{T}}_\eta \cdot [\mathbf{D}_{\bar{T}}]_{1:\eta,\eta+1:N} \right]. \tag{149}$$

[61] In the case of Euclidean kernels, the Nyström "approximation" is actually an exact completion if the entries of the required minors are error-free.

Finally, invoke the functional relation existing between the EDM and the Euclidean kernel matrix defined in equation (17), namely

$$\breve{\mathbf{D}}_{\bar{T}} = \mathbf{1}_N \cdot \mathrm{diag}(\breve{\mathbf{K}})^{\mathrm{T}} + \mathrm{diag}(\breve{\mathbf{K}}) \cdot \mathbf{1}_N^{\mathrm{T}} - 2 \cdot \breve{\mathbf{K}}. \tag{150}$$

Equation (150) yields a complete set of distances associated with $\breve{\mathbf{K}}$, such that any missing entries of $[\mathbf{D}]_{\eta+1:N,\eta+1:N}$ can be replaced by corresponding entries from $\breve{\mathbf{D}}_{\bar{T}}$.

At this point, let us emphasize that $[\mathbf{D}_{\bar{T}}]_{1:\eta,1:\eta}$ contains the distances amongst anchors and consequently $[\mathbf{K}]_{1:\eta,1:\eta}$, $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ are all *constant*, such that $\breve{\mathbf{K}}$ can be updated very efficiently. Furthermore, the elements of $[\mathbf{D}_{\bar{T}}]_{1:\eta,\eta+1:N}$ are the distances from anchors to targets, and therefore constitute the least (reasonable) amount of information required by tracking applications, such that this "completion" procedure can always[62] be applied.

Notice that also in the extreme case where $[\mathbf{D}]_{\eta+1:N,\eta+1:N} = \mathbf{0}_{N-\eta}$, then $[\mathbf{X}]_{\eta+1:N,1:\eta}$ can be recovered from the eigenspace of $[\breve{\mathbf{K}}]_{1:\eta,1:\eta}$ [148]. Indeed, let $[\breve{\mathbf{K}}]_{1:\eta,1:\eta} = \mathbf{Q} \cdot \mathbf{H} \cdot \mathbf{Q}^{\mathrm{T}}$ be the eigen-decomposition of $[\breve{\mathbf{K}}]_{1:\eta,1:\eta}$. From equation (77) it follows that $[\mathbf{X}]_{1:\eta,1:\eta} = \mathbf{Q} \cdot \mathbf{H}^{\odot\frac{1}{2}}$, and because $[\breve{\mathbf{K}}]_{1:\eta,\eta+1:N} = [\mathbf{X}]_{1:\eta,1:\eta} \cdot [\mathbf{X}]_{\eta+1:N,1:\eta}^{\mathrm{T}}$ then the estimate of $[\mathbf{X}]_{\eta+1:N,1:\eta}$ reduces to

$$[\mathbf{X}]_{\eta+1:N,1:\eta} = [\mathbf{X}]_{1:\eta,1:\eta}^{-\mathrm{T}} \cdot [\breve{\mathbf{K}}]_{1:\eta,\eta+1:N} = \mathbf{Q} \cdot \mathbf{H}^{\odot\frac{1}{2}} \cdot [\breve{\mathbf{K}}]_{1:\eta,\eta+1:N}. \tag{151}$$

Therefore, when an incomplete EDM is observed the aforementioned steps can be followed to complete $\mathbf{D}_{\bar{T}}$ before constructing the *dynamic* MDS kernel $\mathbf{K}^*$ as described by equations (18) and (84). Figure 22 shows that by completing the eventual missing entries of $[\mathbf{D}_{\bar{T}}]_{\eta+1:N,\eta+1:N}$ using the Nyström approximation described above it is possible to exploit all the information available in $\mathbf{D}_{\bar{T}}$ to increase the accuracy of the estimates.

While the Nyström approach discussed above well suits tracking scenarios in which the rows of $\mathbf{D}_{\bar{T}}$ corresponding to the anchor-to-target measurements are assumed always known, in presence of mesh networks where the observed EDM matrices are punctured randomly, the method cannot be applied as it is.

[62] As discussed in the following, even the case of sparse incomplete EDMs in which *none* of the rows of $\mathbf{D}_{\bar{T}}$ is complete can also be dealt with by combining the Nystöm solution with standard completion algorithms applied to a restricted subset of η rows of $\mathbf{D}$ [170].

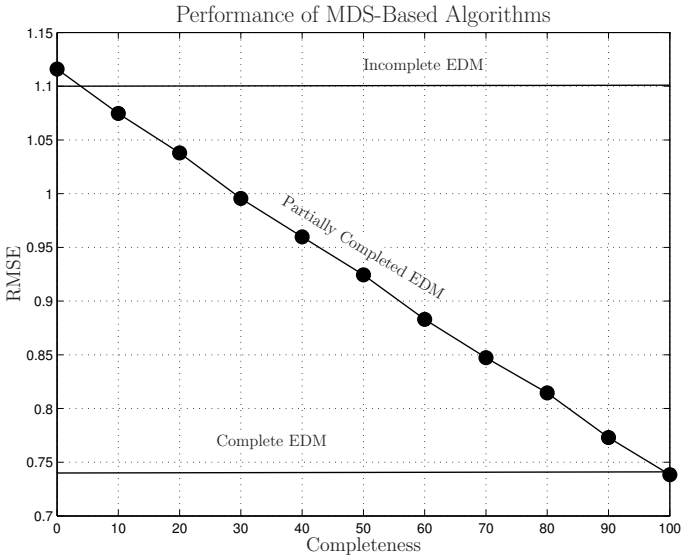


Fig 22. Performance of MDS-based algorithms as function of the completion of $[\mathbf{D}]_{\eta+1:N,\eta+1:N}$. The scenario includes 4 anchor nodes placed at the edges of a 10×10 square in a $\eta = 2$ dimensional space with 21 targets randomly placed in the space and the entries of $\mathbf{D}$ perturbed by $\sim p_{\textsf{Gauss}}(r; 0, 1)$. Revised form [12].

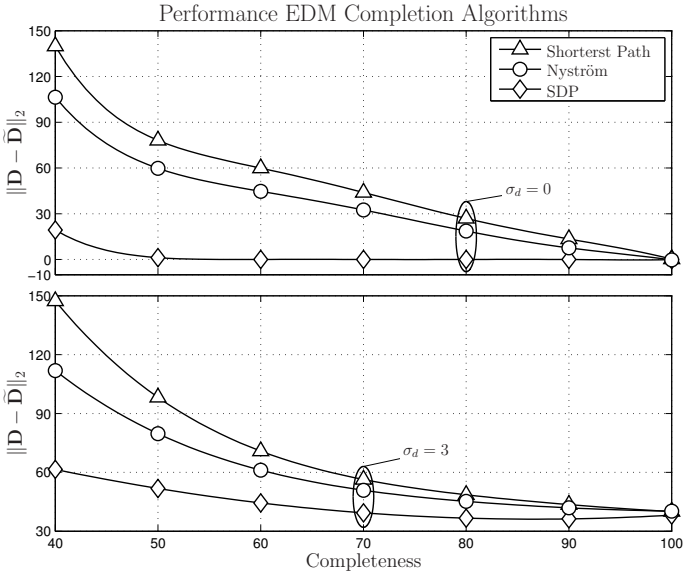


Fig 23. Comparison EDM completion algorithms for a localization scenario with 4 anchor nodes placed at the corners of a 10×10 space ($\eta = 2$) and 10 targets randomly located within the *convex-hull* defined by the anchors. Revised form [12].

However, under the condition that the network is percolated (as required by Graph-based Completion) this problem can be overcome by first completing $\eta + 1$ rows of $\mathbf{D}$ using Dijkstra's algorithm and only then $\mathbf{K}$ is approximated using the Nyström method above.

The performance of this *hybrid* Nyström completion solution is compared in Figure 23 against Dijkstra' shortest path [170] solution and the semidefinite programming approach suggested in [50]. Clearly the computational complexity required by the three completion techniques under consideration is very different. Indeed while the Nyström based solution only requires the usage of Dijkstra's algorithm on $\eta + 1$ rows of $\breve{\mathbf{D}}$ and few matrix multiplications, the shortest path needs to complete all the N rows of the matrix while the SDP method needs to find the solution of a complex constrained optimization.

As expected, Figure 23 confirms, especially in the case of low levels of completion of $\mathbf{D}$, the superiority of the SDP approach in presence of perfect as well as corrupted distance information. However, the same figure also shows that the gain is performance for the SDP method is heavily attenuated in presence of noisy EDMs. The localization accuracy achieved by the Nyström method and the SDP solution proposed in [51] are compared in Figure 24.

Once more the SDP approach is shown to outperform the proposed method reaching the theoretical limit expressed by the CRLB formulated under the assumption of *full* (deterministic) knowledge on the location of the targets.[63]

However, in light of the low complexity of the Nyström method and, similarly to the philosophy adopted in the SDP[64], it is possible to use equation (151) as the initial guess for an optimization algorithm. The resulting hybrid Initialization scheme has been tested on the SMACOF algorithm presented in Section 3.1.2.[65]

Figure 24 compared the performance of the algorithms for different meshness ratio and the entries of $\tilde{\mathbf{D}}$ subject to a perturbation $\sim p_{\text{Gauss}}(r; 0, \sigma_d)$ with $\sigma_d = 1$. The results show that for almost all the meshness values considered, the proposed hybrid initialization algorithms brings the performance of low-

[63] Closed form expressions for the FIM necessary to compute the CRLB for network localization problems be easily found in the literature, *e.g.* [69].

[64] Notice that the solution to the localization problem through an SDP formulation, *e.g.* the algorithms proposed in [50, 51], always assume a *refinement* stage which is usually performed using a non-linear least square (NLS) algorithm.

[65] In the following the SMACOF algorithm initialized with the proposed hybrid Nyström based solution is indicated by SMACOF*.

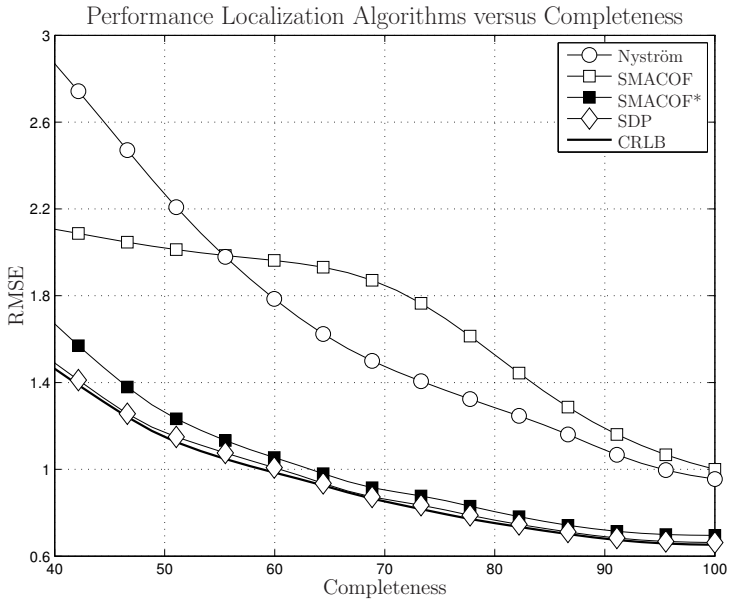


Fig 24. Performance of MDS-based algorithms as function of the completion of $[\mathbf{D}]_{\eta+1:N,\eta+1:N}$. The scenario includes 4 anchor nodes placed at the edges of a 10×10 square in a $\eta = 2$ dimensional space with 10 targets randomly placed in the space and the entries of $\mathbf{D}$ perturbed by $\sim p_{\text{Gauss}}(r; 0, 1)$. Revised form [19].

complexity algorithm, such as the SMACOF approach, close to that of far more sophisticated alternatives such as the SDP in [51].

Moreover the solution almost always approaches the CRLB and only for very low values of completeness the SMACOF* diverges slightly.

5.2 Heterogeneous scenarios

In the following the integration of heterogeneous information within the C-MDS framework is dealt with. In particular it is shown that using the C-MDS technique to solve the edges of graph-associated network topology rather than the points' coordinates, allows to jointly process distance and AOA information in the optimization.

This revised C-MDS technique, labeled super multidimensional scaling (S-MDS) algorithm, exhibits the potential to be robust to erroneous angle

information [23]. However, one drawback of the original S-MDS approach is that it requires the distances between all pairs of points, as well as the angles of the triangles formed by all triad of points. Clearly the collection of all such information may be a problem in many multisource localization scenarios of interest, in particular those characterized by networks with meshy topologies of relatively large dimensions, where node-to-node distances and angles can only be measured within vicinities. A second main problem with the S-MDS solution is that it requires the eigen-decomposition of a kernel matrix whose size increases quadratically with the number of nodes in the network.

In the following the S-MDS algorithm is reconsidered and modified to handle the aforementioned problems by incorporating the Nyström approximation and the interval analysis tools respectively.

5.2.1 Partial heterogeneous information

Section 3.1.1 showed that through equation (13) the dissimilarity matrix $\mathbf{D}$ can be transformed into the inner product matrix $\mathbf{K}$. Moreover, the definition of inner product between points for the ij-th element of $\mathbf{K}$ yields

$$[\mathbf{K}]_{i,j} \triangleq \langle \boldsymbol{x}_i; \boldsymbol{x}_j \rangle = d_i d_j \cos(\theta_{i,j}). \tag{152}$$

It follows that the kernel matrix $\mathbf{K}$ can be written as

$$\mathbf{K} = \boldsymbol{\Theta}_{\mathbf{K}} \circ \left(\mathbf{d}_N \cdot \mathbf{d}_N^{\mathrm{T}} \right), \tag{153}$$

where $\circ$ is the Hadamard product, $\mathbf{d}_N$ contains the distances from the origin for the N points in $\mathbf{X}$ and

$$\boldsymbol{\Theta}_{\mathbf{K}} = \begin{bmatrix} 1 & \cdots & \cos(\theta_{1,N}) \\ \vdots & \ddots & \vdots \\ \cos(\theta_{N,1}) & \cdots & 1 \end{bmatrix}, \tag{154}$$

includes the angles for the M edges in $\mathbf{X}$ as seen from the origin.

Using the definition of inner product matrix given in equation (153) then $\mathbf{X}$ can be recovered from a kernel constructed on the basis of a subset of all the possible range and angle measurements. A by-product of this feature is that error propagation described in equation (150) is largely avoided, since the

Gram kernel is computed (element-by-element), straight from distance and angle estimates.

However, to be of any practical use the kernel $\mathbf{K}$ needs to be centered at a i-th point of $\mathbf{X}$ from where those distances and angles can be measured[66]. Also notice that the resulting $\mathbf{K}_{(i)}$ kernel is equivalent to the one obtained starting from the full set of all mutual distances between the points in $\mathbf{X}$ and computed using equation (13) with $\mathbf{a}$ set to be everywhere zero except in correspondence of the i-th position.[67] Once the new kernel matrix $\mathbf{K}_{(i)}$ is available, $\mathbf{X}$ is recovered applying equation (16) followed by a Procrustes transformation detailed in Appendix 1.

5.2.2 Full heterogeneous information

Although $\mathbf{K}_{(i)}$ allows to process angle and distance information jointly, it only includes only one of all possible subsets of distance and angle measurements amongst the points in $\mathbf{X}$. An extension of the C-MDS that potentially can exploit all mutual distance and angle information was proposed in [23] under the name of S-MDS.

The key idea in the S-MDS algorithm is to solve the cost function in equation (14) for the kernel matrix associated the graph of $\mathbf{X}$. Specifically, let us associate to the network configuration a complete oriented graph $G_{\eta,N}(\mathbf{X}, \vec{\mathbf{V}}, \mathbf{D})$, where $\vec{\mathbf{V}} = \{\vec{\boldsymbol{v}}_{n,m}\}$ is the set of edges with arbitrary, but unique, orientations (vectors). For convenience, the orientation of all edges is chosen so that we may order the M elements of $\vec{\mathbf{V}}$ progressively, *i.e.*, $\vec{\mathbf{V}} = \{\vec{\boldsymbol{v}}_{1,2}, \vec{\boldsymbol{v}}_{1,3}, \cdots, \vec{\boldsymbol{v}}_{N-1,N}\}$.

In order to simplify the notation, let us also relabel each of the mutually exclusive pairs of indexes (n, m), with $n < m$, by a different number i, such that the edge departing from the vertex $\mathbf{x}_n$ to the vertex $\mathbf{x}_m$ can be denoted simply by,

$$\vec{\boldsymbol{v}}_i = (\mathbf{x}_m - \mathbf{x}_n) = [(x_{m,1} - x_{n,1}), \cdots, (x_{m,\eta} - x_{n,\eta})]^{\mathrm{T}}. \tag{155}$$

[66] The matrix $\boldsymbol{\Theta}$ is recovered from the angle measured from the i-th reference point using equation (152).

[67] Centering on the first point results in $\mathbf{J}$ is the full-rank skinny Schoenberg auxiliary matrix defined by [40]

$$\mathbf{J} = \frac{1}{\sqrt{2}} \begin{bmatrix} -\mathbf{1}^{\mathrm{T}} \\ \mathbf{I}_{N-1} \end{bmatrix} \in \mathbb{R}^{N \times N-1}.$$

Next, define the dissimilarity metric $k_{i,j}$ of the i-th and j-th edges as the inner product,

$$\begin{aligned}[\mathbf{K}_{\mathrm{E}}]_{i,j} &\triangleq \langle \vec{\boldsymbol{v}}_i; \vec{\boldsymbol{v}}_j \rangle = \langle (\mathbf{x}_m - \mathbf{x}_n); (\mathbf{x}_q - \mathbf{x}_p) \rangle \\ &= d_{n,m} d_{p,q} \cos(\theta_{i,j}). \end{aligned} \tag{156}$$

where the set of all dissimilarity measures $[\mathbf{K}_{\mathrm{E}}]_{i,j}$, corresponding to all pairs of edges in the graph, can be conveniently assembled into the edge kernel matrix

$$\begin{aligned}\mathbf{K}_{\mathrm{E}} &= \langle [\vec{\boldsymbol{v}}_1, \cdots, \vec{\boldsymbol{v}}_M]; [\vec{\boldsymbol{v}}_1, \cdots, \vec{\boldsymbol{v}}_M]^{\mathrm{T}} \rangle \\ &= \vec{\mathbf{V}} \cdot \vec{\mathbf{V}}^{\mathrm{T}} = \begin{bmatrix} \langle \vec{\boldsymbol{v}}_1; \vec{\boldsymbol{v}}_1 \rangle & \cdots & \langle \vec{\boldsymbol{v}}_1; \vec{\boldsymbol{v}}_M \rangle \\ \vdots & \ddots & \vdots \\ \langle \vec{\boldsymbol{v}}_M; \vec{\boldsymbol{v}}_1 \rangle & \cdots & \langle \vec{\boldsymbol{v}}_M; \vec{\boldsymbol{v}}_M \rangle \end{bmatrix}. \end{aligned} \tag{157}$$

Notice also that $\vec{\mathbf{V}}$ has the structure of a coordinate matrix associated to the M points that define the vectors $\vec{\boldsymbol{v}}_i$, $i = \{1, \cdots, M\}$.

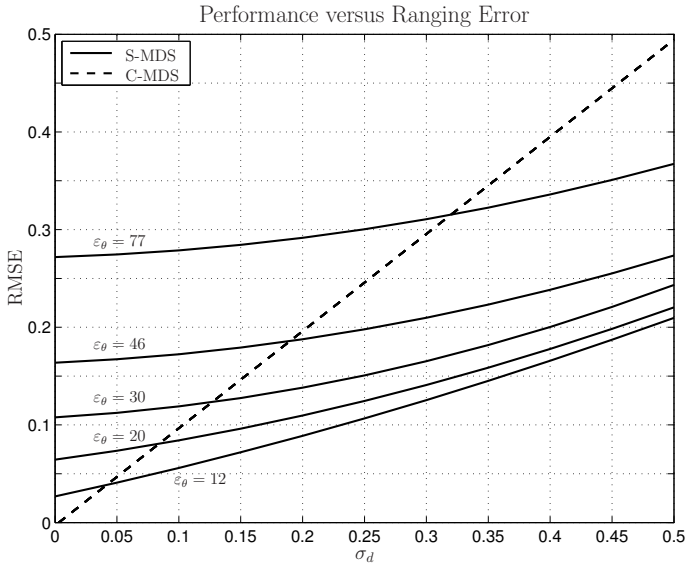
The S-MDS algorithm results directly from the fact that $\mathbf{K}_{\mathrm{E}}$ is a Gram matrix from which it follows that $\vec{\mathbf{V}}$ can be recovered from $\mathbf{K}_{\mathrm{E}}$ using equation (16). The final step of the localization algorithm is to recover $\mathbf{X}$ from the retrieved vector matrix $\vec{\mathbf{V}}$. To this end, consider the following system of linear equations, derived directly from equation (155),

$$\boldsymbol{C} \cdot \mathbf{X} = \vec{\mathbf{V}}, \tag{158}$$

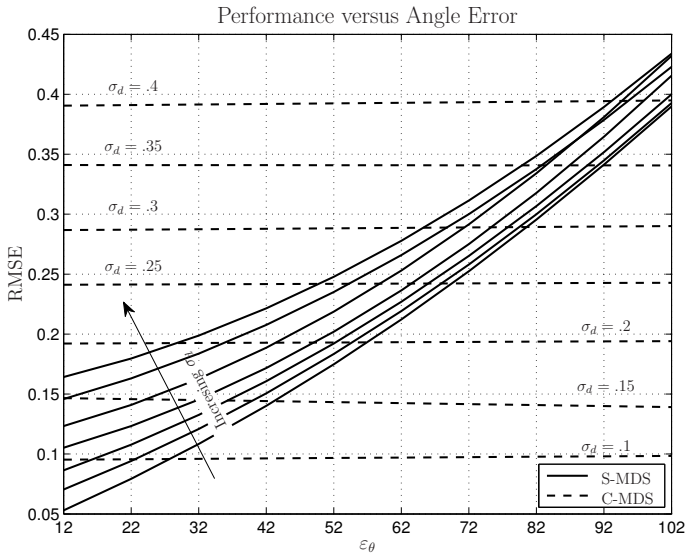
where the coefficient matrix $\boldsymbol{C}$ has the block upper-triangular structure shown below

$$\boldsymbol{C} = \left[\begin{array}{c:cc:cc} \mathbf{1}_{N-1\times 1} & -\mathbf{I}_{N-1\times N-1} & & & \\ \hdashline \mathbf{0}_{N-2\times 1} & \mathbf{1}_{N-2\times 1} & -\mathbf{I}_{N-1\times N-1} & & \\ \hdashline & \ddots & \ddots & \ddots & \\ \hdashline & \mathbf{0}_{1\times N-2} & & 1 & -1 \end{array}\right]. \tag{159}$$

As shown in Figures 25 and 26 one benefit of the decomposition in equation (157) is its robustness to corrupted angle information. Also notice that, as shown in Figure 26(b) during the involutive mapping defined in equation (158) it is possible to *weight* the rows of $\boldsymbol{C}$ and consequently the single entries $\mathbf{D}$ making the S-MDS algorithm robust to the distance erasures problem, which is simply handled by weighting to zero the rows of $\boldsymbol{C}$ corresponding to the missing entries in $\mathbf{D}$.

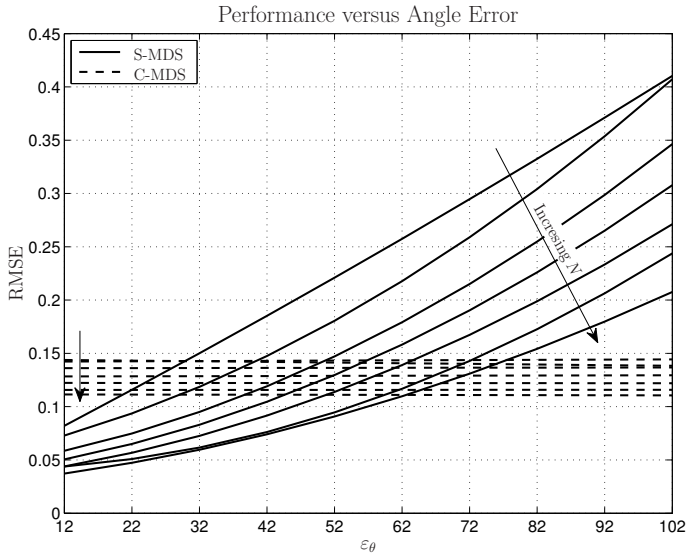


(a) Performance in LOS. RMSE as a function of σ_d.

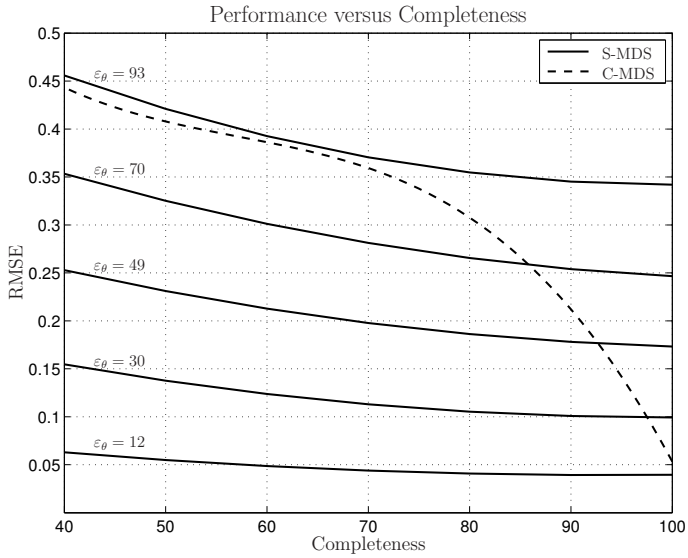


(b) Performance in LOS. RMSE as a function of the angle error ε_θ.

Fig 25. Performance of metric and S-MDS as a function of σ_d and ε_θ for the scenario described in Section 5.3.



(a) Performance in LOS. RMSE as a function of the bias ε_θ and for different values of N.



(b) Performance in LOS. RMSE as a function of the completeness.

Fig 26. Performance of metric and S-MDS as a function of ε_θ and the level of completeness for the range measurements for the scenario described in Section 5.3.

Similarly to the above eventual knowledge on the *confidence* of the range measurements can be incorporated in the estimation process through the matrix $\boldsymbol{C}$.

However, in this form the S-MDS algorithm still has the following limitations:

1. *Increased complexity to eigen-decompose* $\mathbf{K}_{\mathrm{E}}$
2. *All the angles between all the* edges *in* $\vec{\mathbf{V}}$ *are required,*

which are the objective of the following section.

5.2.3 Heterogeneous information with the MDS framework

Section 5.2.2 showed that $\mathbf{K}_{\mathrm{E}} \in \mathbb{R}^{M \times M}$, with M as the number of combinations from the N points in $\mathbf{X}$. The need to eigen-decompose $\mathbf{K}_{\mathrm{E}}$ represents a bottleneck of the S-MDS algorithm when applied to networks whose number of nodes N is sufficiently large.

Although different algorithms could be used for that purpose [143], since $\mathbf{K}_{\mathrm{E}}$ is a Gram matrix, then the numerical approximation to the eigenfunction problem used in the context of kernel-based predictors introduced in Section 5.1.1, namely the Nyström approximation, can be used to solve the aforementioned problem. Indeed, as already seen in Section 5.1 for the incomplete EDM problem, given the kernel matrix $\mathbf{K}_{\mathrm{E}}$ with $\mathrm{rank}(\mathbf{K}_{\mathrm{E}}) = \eta$, the Nyström algorithm approximates the eigen-problem associated to it as the one of a smaller system and then expand back the result [148].

Let B_A be a number of rows selected at random from the kernel matrix and $B_T = M - B_A$ the remaining edges in $G_{\eta,N}(\mathbf{X}, \vec{\mathbf{V}}, \mathbf{D})$. Without loss of generality also assume that those are permuted to the first B_A position of the the matrix, resulting in

$$\mathbf{K}_{\mathrm{E}} \approx \left[\begin{array}{c:c} \boldsymbol{A}_{B_A \times B_A} & \mathbf{T}_{B_A \times B_T} \\ \hdashline \mathbf{T}^{\mathrm{T}}_{B_T \times B_A} & \mathbf{T}^{T}_{B_T \times B_A} \cdot \boldsymbol{A}^{-1}_{B_A \times B_A} \cdot \mathbf{T}_{B_A \times B_T} \end{array}\right]. \tag{160}$$

Let $\boldsymbol{A} = \mathbf{U}_{\boldsymbol{A}} \cdot \boldsymbol{\Lambda}_{\boldsymbol{A}} \cdot \mathbf{U}^{\mathrm{T}}_{\boldsymbol{A}}$ be the eigendecomposition of the first minor of $\mathbf{K}_{\mathrm{E}}$ in equation (160), then equation (16) yields $\vec{\mathbf{V}}_{\boldsymbol{A}} = [\mathbf{U}_{\boldsymbol{A}}]_{B_A \times \eta} \cdot [\boldsymbol{\Lambda}_{\boldsymbol{A}}]^{\odot \frac{1}{2}}_{1:\eta}$, and since $\mathbf{K}_{\mathbf{T}} = \vec{\mathbf{V}}_{\boldsymbol{A}} \cdot \mathbf{T}^{\mathrm{T}}$,

$$\vec{\mathbf{V}}_{\mathbf{T}} = \vec{\mathbf{V}}^{-\mathrm{T}}_{\boldsymbol{A}} \cdot \mathbf{T} = [\boldsymbol{\Lambda}_{\boldsymbol{A}}]^{\odot\left(-\frac{1}{2}\right)}_{\eta \times \eta} \cdot [\mathbf{U}_{\mathbf{A}}]^{\mathrm{T}}_{B_A \times \eta} \cdot \mathbf{T}, \tag{161}$$

where $\vec{\mathbf{V}} = \left[\vec{\mathbf{V}}_{\boldsymbol{A}}^{\mathrm{T}}, \vec{\mathbf{V}}_{\mathbf{T}}^{\mathrm{T}}\right]^{\mathrm{T}}$.

Hence $\vec{\mathbf{V}}_{\mathbf{T}}$ can be recovered from the eigen-decomposition of $\boldsymbol{A}$ only. This approach is particularly convenient in all those cases, such as the multitarget localization scenario here considered, where, $B_A \ll B_T$. Indeed, due to the combinatorial relationship existing between the number of nodes and edges in a graph, when $N_T > N_A$ then $B_T \gg B_A$. It follows that the Nyström approximation of the kernel $\mathbf{K}_{\mathrm{E}}$ only relies on the first B_A rows of $\mathbf{K}_E$. More importantly the edges associated to the targets are computed directly by means of equation (161), namely without involving any decomposition.

A further implication of the Nyström method can be understood expressing the edge kernel $\mathbf{K}_{\mathrm{E}}$ similarly to equation (153), namely

$$\mathbf{K}_{\mathrm{E}} = \boldsymbol{\Theta}_{\mathrm{E}} \circ \left(\mathbf{d}_M \cdot \mathbf{d}_M^{\mathrm{T}}\right), \tag{162}$$

where $\boldsymbol{\Theta}_{\mathrm{E}}$ is the matrix including all the edge-angles as seen from the origin and $\mathbf{d}_M \in \mathbb{R}^{M\times 1}$ the vector containing the corresponding edge length for each one of the M edges in $G_{\eta,N}(\mathbf{X}, \vec{\mathbf{V}}, \mathbf{D})$.

Due to equation (160) than an approximation of $\mathbf{K}_{\mathrm{E}}$ can be constructed on the basis of

$$\boldsymbol{A} = \quad \boldsymbol{\Theta}_{\boldsymbol{A}} \circ \left(\mathbf{d}_{B_A} \cdot \mathbf{d}_{B_A}^{\mathrm{T}}\right), \tag{163}$$

$$\mathbf{T} = \quad \boldsymbol{\Theta}_{\mathbf{T}} \circ \left(\mathbf{d}_{B_T} \cdot \mathbf{d}_{B_T}^{\mathrm{T}}\right), \tag{164}$$

where $\boldsymbol{\Theta}_{\boldsymbol{A}}$ and $\boldsymbol{\Theta}_{\mathbf{T}}$ are respectively the matrix including all the anchor-to-anchor and anchor-to-target angles as seen from the origin while $\mathbf{d}_{B_A} \in \mathbb{R}^{B_A\times 1}$ and $\mathbf{d}_{B_T} \in \mathbb{R}^{B_T\times 1}$ are the vectors containing length fo the corresponding edges.

Since $\mathbf{d}_M = [\mathbf{d}_{B_A}, \mathbf{d}_{B_T}]$ it is evident that the Nyström approximation is capable to retain all range information in $\mathbf{d}_M$ and only affects the angle information included in $\boldsymbol{\Theta}_{\mathrm{E}}$ by neglecting the mutual angles between the targets. This can be particularly convenient in certain LT applications, *e.g.* cellular network scenarios, where it can be expected that a subset of nodes, *e.g.* the base stations, are equipped with the hardware necessary to estimate angle of arrivals.

In light of the above the S-MDS multitarget localization algorithm can be summarized as follows:

Inputs:

- *Anchor's coordinates*
- *Pairwise distance estimates amongst all the nodes*
- *Anchor-to-target pairwise angle estimates.*

Steps:

1. *Construct* $\mathbf{T}$ *using equation* (156)
2. *Estimate the edge vector* $\vec{\mathbf{V}}_{\mathbf{T}}$ *using equation* (161)
3. *Apply the Procrustes transformation on* $\vec{\mathbf{V}}_{\mathbf{T}}$
4. *Map edges into points using equation* (158).

5.2.4 Interval analysis for angle completion

Even though Section 5.2.2 already dealt with the problem of eventual erasures in $\mathbf{d}_M$, the S-MDS approach still requires that all the angles in $\mathbf{\Theta}_{\mathrm{E}}$, or at least the ones in its first B_A rows using the Nyström approximation, are known.

Fortunately, however, in light of the robustness of the S-MDS method to erroneous angle information shown in Figures 25-26 it is possible to complete or even replace the angle matrix $\mathbf{\Theta}_{\mathrm{E}}$ with the following estimate

$$\hat{\mathbf{\Theta}}_{\mathrm{E}} = \left(\hat{\vec{\mathbf{V}}}_{\mathrm{E}} \cdot \hat{\vec{\mathbf{V}}}_{\mathrm{E}}^{\mathrm{T}}\right) \oslash \left(\mathbf{d}_{\mathrm{E}} \cdot \mathbf{d}_{\mathrm{E}}^{\mathrm{T}}\right), \tag{165}$$

where $\hat{\vec{\mathbf{V}}}_{\mathrm{E}} = \mathcal{V}(\hat{\mathbf{X}})$ is the matrix of edges constructed from an initial estimates of $\hat{\mathbf{X}}$, $\oslash$ is the inverse of $\circ$ and $\mathbf{d}_{\mathrm{E}} \in \mathbb{R}_{+}^{M \times 1}$ is defined as

$$\mathbf{d}_{\mathrm{E}} = \left(\mathrm{diag}\left(\hat{\vec{\mathbf{V}}}_{\mathrm{E}} \cdot \hat{\vec{\mathbf{V}}}_{\mathrm{E}}^{\mathrm{T}}\right)\right)^{\odot \frac{1}{2}}, \tag{166}$$

where $\mathrm{diag}(\cdot)$ is a function that returns the diagonal elements of the matrix given as argument.

Provided that $\hat{\mathbf{\Theta}}_{\mathrm{E}}$ is known an approximation for edge kernel matrix is obtained as

$$\hat{\mathbf{K}}_{\mathrm{E}} = \hat{\mathbf{\Theta}}_{\mathrm{E}} \circ \left(\mathbf{d}_M \cdot \mathbf{d}_M^{\mathrm{T}}\right). \tag{167}$$

Clearly there is still the problem to find the appropriate $\hat{\mathbf{\Theta}}_{\mathrm{E}}$. One method that does that on the basis of the set of anchor-to-target distance measurements only is described in the following.

Let $\{d_{1,j}, \ldots, d_{N_A,j}\}$ be the set of distance measurements between the j-th target and the anchor-to-target measurements, then it is known that $[\hat{\mathbf{X}}_C]_j$ can be found jointly solving a set of $i = \{1, \ldots, B_A\}$ constraints [35]. Specifically, in the case of planar configurations[68], namely $\eta = 2$ and denoting with $\hat{\mathbf{x}}_j = [x_j, y_j]$ the j-th target's coordinates, the relation with respect to the i-th anchor can be expressed by

$$\| \mathbf{x}_i - \mathbf{x}_j \| = d_{i,j} \Rightarrow (x_i - x_j)^2 + (y_i - y_j)^2 = d_{i,j}. \tag{168}$$

In presence of noisy range measurements, however, equation (168) needs to be modified. In particular, under the case of non-line of sight (NLoS) conditions, the anchor-to-target equality in equation (168) changes into [35]

$$\| \mathbf{x}_i - \mathbf{x}_j \| \leqslant d_{i,j} \Rightarrow (x_i - x_j)^2 + (y_i - y_j)^2 \leqslant d_{i,j}. \tag{169}$$

It is also known [35] that each one of the i-th circular constraints above can be relaxed and expressed by a set of η linear constraints as

$$x_i - x \leqslant d_{i,j} \quad \text{and} \quad -x_i + x \leqslant d_{i,j}, \tag{170}$$

$$y_i - y \leqslant d_{i,j} \quad \text{and} \quad -y_i + y \leqslant d_{i,j}. \tag{171}$$

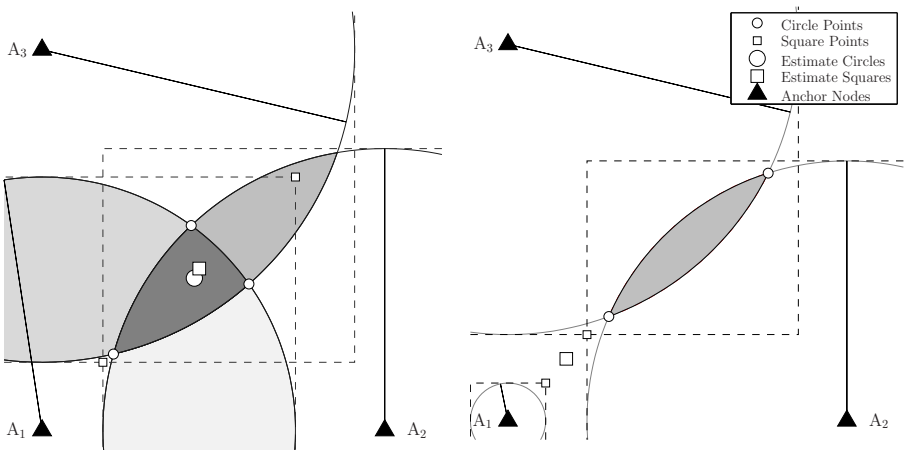
Let $\mathcal{A}^{(i)}$ be the area defined by the i-th constraint relative to the j-th target, then the intersection of the N_A constraints defines a *feasibility* region in the space where the target is ensured to be, namely,[69]

$$\mathbf{x}_j \in \mathcal{A}_{\mathrm{T}} = \bigcap_{i=1}^{N_A} \mathcal{A}^{(i)}. \tag{172}$$

A representation of such region obtained from the set N_A inequalities in equation (169) or similarly the set of linear constraints in equation (170) is provided in Figure 27(a) and 27(b) for different types of errors added to the true anchor-to-target range measurements.

[68] The extension to the case of 3-D scenarios is straightforward.

[69] As Figure 27(a) shows the region given by the union of all the linear constraints is only an upper bound to one obtained using the circular constraints.



(a) Measurements affected by positive biases. (b) Measurements affected by positive and negative biases.

Fig 27. Feasibility region $\mathcal{A}_{\mathrm{T}}$ defined in equation (169) and (170) in presence of positive biases (left) and positive and negative biases (right). Revised form [18].

In absence of any additional information about the range measurements, *e.g* statistical model, each point inside the feasibility region can be assumed as equally likely, implying that an estimate for $\hat{\mathbf{x}}_j$ can be obtained as the center of mass of the region itself. Specifically for the circular constraints such a point can be found by intersecting the radial lines defined by the pair or inequalities [33]. Differently the feasibility region defined by the linear constrains is fully characterized by a *lower* and an *upper* point whose coordinates are [35]

$$\mathbf{x}_\ell = \left[\min_i\{x_i + d_{i,j}\}, \min_i\{y_i + d_{i,j}\}\right], \tag{173}$$

$$\mathbf{x}_u = \left[\max_i\{x_i - d_{i,j}\}, \max_i\{y_i - d_{i,j}\}\right]. \tag{174}$$

It follows that the center of mass for the linear constraints is simply the average of the aforementioned points.[70]

To constrain $\hat{\mathbf{x}}$ to the feasibility region $\mathcal{A}_{\mathrm{T}}$ already found wide application in constrained-optimization solution to the localization problem under NLOS conditions [35, 171]. However, those optimization techniques are usually computationally demanding when compared to method here proposed. Eventual

[70] As shown on the right hand side of Figure 27 differently from the system of linear constraints that always return an estimated $\hat{\mathbf{x}}_j$, the circular constraints only work when in presence of biased or perfect measurements [35].

extensions of the algorithm to dynamic scenarios are possible by combining the approach above with the technique suggested in [172].

5.3 Performance evaluation and comparisons

In the following the performance of the S-MDS technique is compared against the SMACOF algorithm [37] and the C-MDS based method described in Section 5.2.1. The comparison is carried out under the case of mere range information, using the solution suggested in Section 5.2.4, as well as in heterogeneous scenarios where both range and angle observations between the nodes are measured.

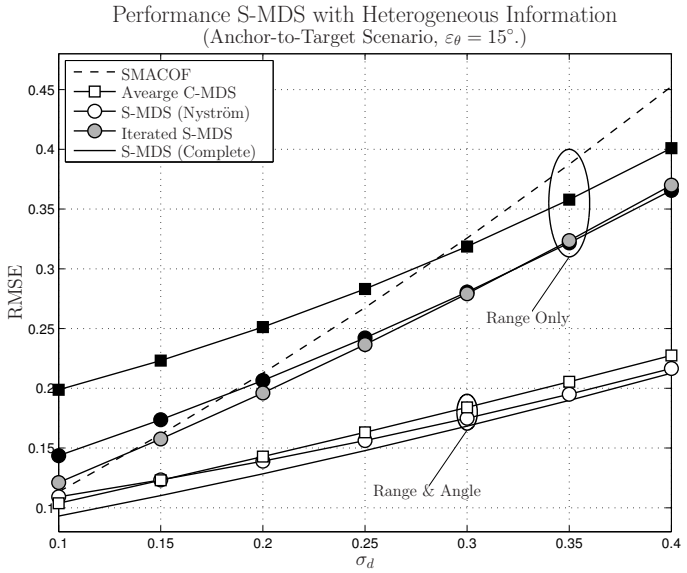
The simulations are designed to provide insight on how much accuracy can be gained by utilizing angle information and how robust the edge-kernel formulation is compared to the other alternatives. The scenario chosen for this comparison consists of 4 anchor nodes placed at edges of a 2×2 square in a $\eta = 2$ dimensional space and 5 targets uniformly displaced in the space centered on the area defined by the anchor's positions. All metric units are normalized and, therefore, shall be omitted henceforth. Although the S-MDS algorithm allows to consider range and angle measurements from each pair of sensors, in practice it can not be expected to collect this information from each single device in the network.

On the contrary it is quite reasonable to assume that only a limited subset of nodes is able to provide both measurements. For this reason below we will restrict ourselves to a cellular network-like scenario in which only a limited number (N_A) of nodes, the anchor nodes, can measure angle and range with all the users in their surroundings.

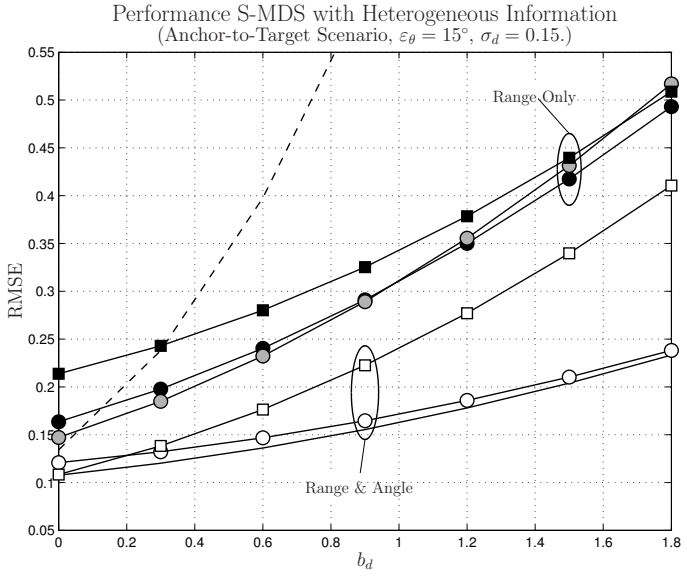
Range estimates, the noisy measurements of the distance between each pair of sensors, are modeled as a Gaussian-distributed random variables with mean given by the true distance and a standard deviation σ_d related to the ranging error affecting its measurement. NLOS conditions are modeled by adding to the mean of the aforementioned Gaussian random variable a sample from $\mathcal{U}(0, b_d)$. In turn, angle estimation errors are modeled as random processes distributed according to the Tikhonov probability density function (pdf) [173]

$$p_{\text{Tikhonov}}(\theta; \rho) = \frac{1}{2\pi I_0(\alpha)} \cdot e^{\rho \cos(\theta)}, \quad \theta \in [-\pi, \pi],\ \rho \geqslant 0. \tag{175}$$

The parameter ρ controls the shape of this pdf, such that $p_{\text{T}}(\theta; \rho)$ tends to a uniform distribution for $\rho \to 0$, and to a Dirac delta at 0 when $\rho \to \infty$.



(a) Performance in LOS. RMSE as a function of σ_d.



(b) Performance in NLOS. RMSE as a function of the bias b_d.

Fig 28. Performance of metric and S-MDS as a function of σ_d and b_d, for range only and range-bearing measurements. Revised form [18].

For a given ρ, we shall define the angle error ε_θ as the average absolute values added to the true θ. The range and angle estimates perturbed according to the Gaussian and Tikhonov distributions are then fed to the S-MDS localization algorithm described above.

The performance of the algorithms are compared in Figure 28 for a LOS and NLOS conditions. The plots reveal that the S-MDS algorithm implemented as discussed in Section 5.2.3 has better accuracy than both the SMACOF algorithm[71] and the average of the method proposed in Section 5.2.1 computed at each one of the N_A anchor nodes. This advantage of the S-MDS solution becomes particularly evident in presence of NLOS measurements. Moreover, similarly to iterated Bayesian solutions, *e.g.* the iterated KF [30], in the case of range only observations an additional improvement in performance can be achieved by iterating the solution, namely updating $\hat{\mathbf{\Theta}}_C$ on the basis of the first solution and performing the S-MDS algorithm described in Section 5.2.3 once more on the same observed range measurements. Concluding, as Figure 26(b) shows a further advance of the S-MDS algorithm is its robustness to data erasures in $\mathbf{d}_M$.

5.4 Summary and discussions

This Chapter extends the application scenario of the MDS-based method to incomplete Euclidean distance matrices and/or heterogeneous information, namely TOA and AOA. To do so, we suggested an efficient Nyström-inspired method to compensate for data erasures in $\mathbf{D}$ and showed how this algebraic solution can be used in conjunction with low-complexity optimization solution to reach the theoretical limit represented by the CRLB.

The problem of using heterogeneous information within the same framework was handled by exploiting the definition of inner product matrix for the kernel matrix $\mathbf{K}$ and solving instead of the targets coordinates directly, the edges associated to the graph resulting from the network configuration. It was shown that is so doing it is possible to compensate biased range information as well as weight the single distance measures.

[71] The algorithm is initialized with the centroids used to compute $\hat{\mathbf{\Theta}}_C$.

6 The passive MTT problem: incomplete and heterogeneous scenarios

This chapter deals with the MTT problem in passive scenarios. To this end the RFS formulation is used to cast the problem into a Bayesian framework which is subsequently solved using the CPHD algorithm detailed in Section 3.2.1. Within this framework the chapter extend the results shown in [20] by incorporating the squared-root implementation of the Cubature Kalman Filter (S-CKF) recently proposed in [84] into the GM-CPHD recursion, resulting in better robustness and numerical stability of the algorithm. Subsequently it is show that the processing requirements for the aforementioned algorithm can be substantially reduced without compromising performance through the *adaptive* elliptical gating strategy here proposed.

The solution is tested on both a multisensor MTT scenario in which at each time $\bar{k}$ a variable number of range measurements are simultaneously observed at several sensor nodes and in a single sensor setting with range bearing observations. The results show that the adaptive gate applied to the proposed S-CKF-GM-CPHD filter offers lower complexity than the non-gated GM-CPHD filter and better robustness than the standard elliptical gating solution.

6.1 The STT problem

As seen in Section 3.1.3, in the classical Bayesian state space formulation of the STT problem, given the initial confidence on the state vector in the form of $p_0(\mathbf{x}_0)$, the evolution of the system is inferred from real time information (measurements) and prior knowledge in the form of a system model such as

$$\boldsymbol{x}_{\bar{k}} = f(\boldsymbol{x}_{\bar{k}-1}) + \boldsymbol{q}_{\bar{k}-1} \quad \text{(Dynamic Model)} \tag{176a}$$

$$\boldsymbol{z}_{\bar{k}} = g(\boldsymbol{x}_{\bar{k}}) + \boldsymbol{r}_{\bar{k}}, \quad \text{(Measurement Model)} \tag{176b}$$

where $\boldsymbol{q}$ and $\boldsymbol{r}$ denote the process and measurement Gaussian noise respectively.[72]

[72] In the following we assume statistical independence amongst $\boldsymbol{q}$ and $\boldsymbol{r}$. Also, although not considered in this chapter, the system model and the following algorithms can be easily modified to describe the more general case of non-additive Gaussian noise.

The evolution of the state vector $\boldsymbol{x}_{\bar{k}} \in \mathbb{R}^{[n_{\boldsymbol{x}} \times 1]}$ at the time $\bar{k}$ is first *predicted* using equation (176) and subsequently *corrected* on the basis of the observation $\boldsymbol{z}_{\bar{k}} \in \mathbb{R}^{[n_{\boldsymbol{z}} \times 1]}$, according to the measurement model in equation (176b).

While the Bayesian solution [30] of the aforementioned problem is general, it requires the propagation in time of the full *posterior* pdf of the state vector, which is often intractable. In practice, therefore, this problem is often solved through KF presented in Section 3.1.3, in which the posterior pdf is approximated by a Gaussian model, such that only its first two moments (mean and co-variance) need be propagated. The Gaussian approximation employed in the Kalman filter also yields typically easy updating of the system model when the functions f and g are linear.

It is well known that in presence of linear Gaussian systems the KF solution is optimal. But it is also common to find sub-optimal utilizations of *linearized* versions of the KF, *e.g.* the EKF, beyond linear systems. Although the EKF works reasonably well for somewhat non-linear systems, it requires the computation of a Jacobian matrix in the approximation of the functions f and g, which is often a source of numerical problems, not to mention that there are cases (when f and g are not differentiable) in which the EKF can only be used if a suitable approximation of the Jacobian can be found.

To mitigate the latter problem, under the principle that *it is easier to approximate a pdf than an arbitrary nonlinear function* [83], the UKF extends the KF by employing the unscented transform (UT) whereby a set of deterministically chosen points, also known as *sigma* points, is used to capture high-order moment-matching conditions of the pdf under investigation, which may not be necessarily Gaussian. Variations of the UKF employing different sets of sigma points can be found in the literature [83, 174, 175], and a square-root version of the UKF, namely the square-root unscented Kalman filter (S-UKF), was also proposed [176].

6.1.1 The CKF framework

A different perspective, unifying all the approaches based on the Gaussian approximation of the filtering distribution, *e.g.* EKF and the UKF solution, is provided by the Gaussian moment matching algorithm [177]. This method, also known as Gaussian filter, simply matches the moments of the posterior pdf,

assumed Gaussian, to the values obtained through the following recursion:
Step 1 - Prediction

$$\mathbf{m}_{\bar{k}|\bar{k}-1} = \int_{\mathbb{R}^{n_{\boldsymbol{x}}}} f(\boldsymbol{x}_{\bar{k}-1}) \cdot \mathcal{N}(\boldsymbol{x}_{\bar{k}-1}; \boldsymbol{m}_{\bar{k}-1}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}-1}) \cdot d\boldsymbol{x}_{\bar{k}-1}, \tag{177}$$

$$\begin{aligned}\mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1} = &\int_{\mathbb{R}^{n_{\boldsymbol{x}}}} \left(f(\boldsymbol{x}_{\bar{k}-1}) - \mathbf{m}_{\bar{k}|\bar{k}-1}\right) \cdot \left(f(\boldsymbol{x}_{\bar{k}-1}) - \mathbf{m}_{\bar{k}|\bar{k}-1}\right)^{\mathrm{T}} \\ &\cdot \mathcal{N}(\boldsymbol{x}_{\bar{k}-1}; \mathbf{m}_{\bar{k}-1}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}-1}) \cdot d\boldsymbol{x}_{\bar{k}-1} + \boldsymbol{Q}_{\bar{k}-1}.\end{aligned} \tag{178}$$

Step 2 - Update

$$\mathbf{m}_{\bar{k}|\bar{k}} = \boldsymbol{m}_{\bar{k}|\bar{k}-1} + \boldsymbol{K}_{\bar{k}} \cdot (\boldsymbol{z}_{\bar{k}} - \boldsymbol{z}_{\bar{k}|\bar{k}-1}), \tag{179}$$

$$\mathbf{P}_{\boldsymbol{x},\bar{k}||\bar{k}} = \mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1} - \boldsymbol{K}_{\bar{k}} \cdot \mathbf{S}_{\bar{k}|\bar{k}-1} \cdot \boldsymbol{K}_{\bar{k}}^{\mathrm{T}}, \tag{180}$$

$$\boldsymbol{z}_{\bar{k}|\bar{k}-1} = \int_{\mathbb{R}^{n_{\boldsymbol{x}}}} g(\boldsymbol{x}_{\bar{k}}) \cdot \mathcal{N}(\boldsymbol{x}_{\bar{k}}; \mathbf{m}_{\bar{k}|\bar{k}-1}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}) d\boldsymbol{x}_{\bar{k}}, \tag{181}$$

$$\begin{aligned}\mathbf{S}_{\bar{k}} = &\int_{\mathbb{R}^{n_{\boldsymbol{x}}}} \left(g(\boldsymbol{x}_{\bar{k}}) - \boldsymbol{z}_{\bar{k}|\bar{k}-1}\right) \cdot \left(g(\boldsymbol{x}_{\bar{k}}) - \boldsymbol{z}_{\bar{k}|\bar{k}-1}\right)^{\mathrm{T}} \\ &\cdot \mathcal{N}(\boldsymbol{x}_{\bar{k}}; \mathbf{m}_{\bar{k}|\bar{k}-1}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}) \cdot d\boldsymbol{x}_{\bar{k}} + \boldsymbol{R}_{\bar{k}},\end{aligned} \tag{182}$$

$$\begin{aligned}\mathbf{P}_{\boldsymbol{xz},\bar{k}} = &\int_{\mathbb{R}^{n_{\boldsymbol{x}}}} \left(\boldsymbol{x}_{\bar{k}} - \mathbf{m}_{\bar{k}|\bar{k}-1}\right) \cdot \left(g(\boldsymbol{x}_{\bar{k}}) - \boldsymbol{z}_{\bar{k}|\bar{k}-1}\right)^{\mathrm{T}} \\ &\cdot \mathcal{N}(\boldsymbol{x}_{\bar{k}}; \mathbf{m}_{\bar{k}|\bar{k}-1}, \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}) \cdot d\boldsymbol{x}_{\bar{k}},\end{aligned} \tag{183}$$

where $\mathbf{S}_{\bar{k}}$ is the covariance of the innovation term $\boldsymbol{\epsilon}_{\bar{k}} = \boldsymbol{z}_{\bar{k}} - \boldsymbol{z}_{\bar{k}|\bar{k}-1}$, and

$$\boldsymbol{K}_{\bar{k}} = \mathbf{P}_{\boldsymbol{xz},\bar{k}} \cdot \mathbf{S}_{\bar{k}}^{-1}, \tag{184}$$

is the *gain* matrix.[73]

From equation (181)-(183) it is evident that the fundamental challenge of nonlinear Bayesian filtering problems is the *efficient* computation of multidimensional integrals required to obtain the predicted measurement and the error covariance matrices. With this in mind, in [84] it was shown that under the assumption of Gaussian conditional densities, the integrals in the Bayesian recursion are all of the form

$$\mathcal{I}(f) = \int_{\mathbb{R}^{n_{\boldsymbol{x}}}} \zeta(\boldsymbol{x}) \cdot \exp\left(-\boldsymbol{x}^{\mathrm{T}\cdot\boldsymbol{x}}\right) d\boldsymbol{x}, \tag{185}$$

where the function $\zeta(\cdot)$ in the context hereby is either $f(\cdot)$ or $g(\cdot)$, normalized by $(2\pi)^{|\boldsymbol{x}|/2} \cdot \mathbf{P}_{\boldsymbol{x}}$, in which $|\cdot|$ denotes the cardinality operator.

[73] Notice that in case of LG systems the filter is equivalent to the KF described in Section 3.1.3.

It was also shown in [84] that integrals in the form shown in equation (185) can be accurately approximated by applying a third-degree spherical-radial cubature rule, which yields,

$$\mathcal{I}(f) \approx \sum_{i=1}^{N} \varpi^{(i)} \zeta(\boldsymbol{\varepsilon}^{(i)}), \tag{186}$$

with

$$\boldsymbol{\varepsilon}^{(i)} = \sqrt{\frac{N_c}{2}} \cdot [\mathbf{I}, -\mathbf{I}], \tag{187}$$

$$\varpi^{(i)} = 1/N_c, \tag{188}$$

where $\{\varpi^{(i)}, \boldsymbol{\varepsilon}^{(i)}\}, (\forall i = 1, \ldots, N_c)$, are the set of *cubature* points and $N_c \triangleq 2 \cdot n_{\boldsymbol{x}}$.

Although the UKF and the CKF filers both use sets of deterministic points to capture the higher moments of the posterior pdf, the sets differ both in values and numbers.[74] More importantly, the set of cubature points used in the CKF is designed under a Gaussian assumption and with the goal of accurately approximating the *integrals* of the Bayesian recursion (which include the posterior pdf in the integrand), while the sigma points of the UKF are computed with aim at best approximating the *posterior pdf itself* (or its moments, to be precise), without any specific assumption on the latter.

In addition to this philosophical distinction, the structure of the CKF allows for proper square-root version of the filter to be implemented, with clear advantages in term of numerical stability [84], since the cubature points are all positive. In contrast, only a *pseudo* square-root version of the UKF can be generally implemented.[75]

6.2 The S-CKF-GM-CPHD filter for MTT

Besides the issues surrounding the implementation and propagation of the PHD intensity function discussed above, another problem with the original PHD filter [116–118] is that the cardinality is propagated in time through "*a single parameter,* [...] *effectively approximat*[ing] *the cardinality distribution* [by]

[74] While the CKF uses N_c cubature points, generally the UKF uses $N_c + 1$ sigma points [83].
[75] As mentioned in [84], the S-UKF proposed in [176] is only a *pseudo* square-root filter, since it requires a correction step to ensure that the propagated square-root error covariance matrix is positive semidefinite.

a Poisson distribution with a matching mean", such that the variance of the cardinality estimate is directly proportional to the number of targets [119].

To overcome this weakness, Mahler proposed in [135] to propagate in time the cardinality distribution[76] $p_{\mathrm{C},\bar{k}}(n)$, along with the intensity function $\nu_{\bar{k}|\bar{k}}$, under the assumptions that: a) target-generated measurements are independent of each other; b) birth and survival RFS are independent of each other; c) prior and predicted multitarget RFS are i.i.d. cluster processes [178, Def. 6.3.I, pp. 176]; and d) clutter measurements are i.i.d. cluster processes.[77] In the resulting CPHD filter, both $\nu_{\bar{k}|\bar{k}}$ and $p_{\mathrm{C},\bar{k}}$ are propagated in time, such that the cardinality $n_{\bar{k}|\bar{k}}$ can be estimated more efficiently than in equation (66), specifically

$$n_{\bar{k}|\bar{k}}(\boldsymbol{X}) \triangleq \sum_{n=1}^{\infty} n \cdot p_{\mathrm{C},\bar{k}|\bar{k}}(n|\boldsymbol{Z}_{\bar{k}}). \tag{189}$$

Alternatively, a more stable estimate of $n_{\bar{k}|\bar{k}}$ can also be obtained by means of a MAP estimator[78], namely

$$n_{\bar{k}|\bar{k}}(\boldsymbol{X}) \triangleq \arg\sup_{n_{\bar{k}|\bar{k}}} \; p_{\mathrm{C},\bar{k}|\bar{k}}(n|\boldsymbol{Z}_{\bar{k}}), \tag{190}$$

which performs better at low signal-to-noise ratio (SNR) conditions [135].

It was shown in [119] that under the assumption of linear Gaussian systems with state independent survival and detection probabilities – specifically $p_{\mathrm{S},\bar{k}}(\xi) = p_{\mathrm{S},\bar{k}}$ and $p_{\mathrm{D},\bar{k}}(\boldsymbol{x}) = p_{\mathrm{D},\bar{k}}$ – the distributions of the RFS describing the detected/surviving targets $\boldsymbol{S}_{\bar{k}|\bar{k}-1}(\xi)$ are *in fact* GM's. If in addition the intensity of the birth RFS $\boldsymbol{\Gamma}_{\bar{k}}$ is modeled as a GM, then, referring to equation (59), it follows that the posterior intensity of PHD is also a GM. Furthermore, in the linear case, the moments of the GM components can be easily obtained form the system model, the observations and the previous estimates.

In the general case of non-linear systems, however, the moments of the GM components cannot be computed directly (see Section 6.1.1), and instead must be approximated. To do so, one approach would be to linearize the system model, which would lead to an EKF-GM-CPHD filter. Another approach would

[76] An interesting physical interpretation of both the PHD and the CPHD recursions in terms of the *bin-occupancy* filter is provided in [137].

[77] This amounts to relaxing the condition on the cardinality distribution so that it can be arbitrary but with i.i.d. elements.

[78] The cardinality estimates shown in Section 6.4 are obtained using equation (190).

be utilized the unscented transform, which would lead to an UKF-GM-CPHD filter. In light of the discussion in Section 6.1.1, however, and given that we are dealing with Gaussian Mixtures, it is clearly advantageous to instead employ the S-CKF filter. In other words, in what follows a detailed description of the GM-CPHD recursion in which the GM components propagate in time via the S-CKF method will be proposed.

Suffice it to describe the recursion of the $\bar{k}-1$-th to the $\bar{k}$-th instance.[79] Let us therefore assume that the cardinality distribution $p_{\mathrm{C},\bar{k}-1}$, the weights $\omega_{\bar{k}-1}^{(j)}$, the means $\mathbf{m}_{\bar{k}-1}^{(j)}$ and the matrices $\mathbf{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)}$ are known, such that the GM representation of the PHD intensity function is given by

$$\nu_{\bar{k}-1}(\boldsymbol{x}) = \sum_{j=1}^{J_{\bar{k}-1}} \omega_{\bar{k}-1}^{(j)} \cdot \mathcal{N}\left(\boldsymbol{x}; \mathbf{m}_{\bar{k}-1}^{(j)}, \mathbf{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)}\right), \tag{191}$$

where the matrix $\mathbf{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)}$ is obtained from the Cholesky factorization of the state covariance matrix $\mathbf{P}_{\boldsymbol{x},\bar{k}-1}^{(j)}$, *i.e.*, $\mathbf{P}_{\boldsymbol{x},\bar{k}-1}^{(j)} = \mathbf{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)} \cdot \left(\mathbf{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)}\right)^{\mathrm{T}}$.

Then the CPHD recursion for the general nonlinear system model represented in equation (176a) is described by the following two steps.

Step 1 - Predition:

$$p_{\mathrm{C},\bar{k}|\bar{k}-1}(n) = \sum_{j=0}^{n} p_{\Gamma,k}(n-j) \sum_{\ell=j}^{\infty} C_j^{\ell} p_{\mathrm{C},\bar{k}-1}(\ell) p_{\mathrm{S},\bar{k}}^{j} (1-p_{\mathrm{S},\bar{k}})^{\ell-j}, \tag{192}$$

$$\nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}) = \nu_{\mathrm{S},\bar{k}|\bar{k}-1}(\boldsymbol{x}) + \nu_{\Gamma,\bar{k}}(\boldsymbol{x}), \tag{193}$$

where $C_j^{\ell} = (\ell!/j!(\ell-j)!)$ is the binomial coefficient[80], and

$$\nu_{\Gamma,\bar{k}}(\boldsymbol{x}) = \sum_{j=1}^{J_{\Gamma,\bar{k}}} \omega_{\Gamma,\bar{k}}^{(j)} \cdot \mathcal{N} \cdot \left(\boldsymbol{x}; \mathbf{m}_{\Gamma,\bar{k}}^{(j)}, \mathbf{\Sigma}_{\Gamma,\bar{k}}^{(j)}\right), \tag{194}$$

$$\nu_{\mathrm{S},\bar{k}|\bar{k}-1}(\boldsymbol{x}) = p_{\mathrm{S},\bar{k}} \sum_{j=1}^{J_{\bar{k}-1}} \omega_{\bar{k}-1}^{(j)} \cdot \mathcal{N}\left(\boldsymbol{x}; \mathbf{m}_{\mathrm{S},\bar{k}|\bar{k}-1}^{(j)}, \mathbf{\Sigma}_{\mathrm{S},\bar{k}|\bar{k}-1}^{(j)}\right), \tag{195}$$

in which $\nu_{\Gamma,\bar{k}}$ denotes the GM used to model the intensity function for the birth process, the matrix $\mathbf{\Sigma}_{\Gamma,\bar{k}}^{(j)}$ is obtained by the Cholesky decomposition of

[79] The initialization of the cardinality distribution $p_{\mathrm{C},\bar{k}-1}$, and the GM component parameters $\omega_{\bar{k}-1}^{(j)}$, $\mathbf{m}_{\bar{k}-1}^{(j)}$,$\mathbf{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)}$ depend on the scenario and thus will be given in Section 6.4.

[80] Notice that equation (192) is the probability density for the process corresponding to the sum of the birth and the surviving targets cardinality, the latter rewritten as sum of its moments.

$\mathbf{P}_{\Gamma,\bar{k}}^{(j)} = \mathbf{\Sigma}_{\Gamma,\bar{k}}^{(j)} \cdot \left(\mathbf{\Sigma}_{\Gamma,\bar{k}}^{(j)}\right)^{\mathrm{T}}$ and

$$\mathbf{m}_{\mathrm{S},\bar{k}|\bar{k}-1}^{(j)} = \tfrac{1}{N_c} \sum_{i=1}^{N_c} \boldsymbol{\psi}_{\bar{k}|\bar{k}-1}^{*(ji)}, \tag{196}$$

$$\mathbf{\Sigma}_{\mathrm{S},\bar{k}|\bar{k}-1}^{(j)} = \mathbf{Tria}\left(\left[\mathbf{A}_{\bar{k}|\bar{k}-1}^{*(j)},\ \mathbf{\Sigma}_{\boldsymbol{Q},\bar{k}-1}\right]\right), \tag{197}$$

with

$$\mathbf{A}_{\bar{k}|\bar{k}-1}^{*(j)} = \tfrac{1}{\sqrt{N_c}}\left(\mathbf{\Psi}_{\bar{k}|\bar{k}-1}^{*(j)} - \mathbf{m}_{\mathrm{S},\bar{k}|\bar{k}-1}^{(j)} \cdot \mathbf{1}_{N_c}^{\mathrm{T}}\right), \tag{198}$$

$$\mathbf{\Psi}_{\bar{k}-1}^{*(j)} = \left[\boldsymbol{\psi}_{\bar{k}-1}^{*(j1)}, \ldots, \boldsymbol{\psi}_{\bar{k}-1}^{*(jN_c)}\right], \tag{199}$$

$$\boldsymbol{\psi}_{\bar{k}|\bar{k}-1}^{*(ji)} = f\left(\boldsymbol{\psi}_{\bar{k}-1}^{(ji)}\right), \tag{200}$$

$$\boldsymbol{\psi}_{\bar{k}-1}^{(ji)} = \mathbf{\Sigma}_{\boldsymbol{x},\bar{k}-1}^{(j)} \varepsilon^{(i)} + \mathbf{m}_{\bar{k}-1}^{(j)}. \tag{201}$$

In the last equations, $\mathbf{Tria}(\cdot)$ denotes the lower triangular component of a general triangularization algorithm (*e.g.* the R matrix of a QR-factorization), $\mathbf{1}_{N_c}$ is a column vector of length N_c containing only ones, $\boldsymbol{\varepsilon}^{(i)}$ is the i-th cubature point defined in equation (187) and $\mathbf{\Sigma}_{\boldsymbol{Q},\bar{k}-1}$ is the square root factor of $\boldsymbol{Q}$.

Let $J_{\bar{k}|\bar{k}-1} \triangleq J_{\Gamma,\bar{k}} + J_{\bar{k}-1}$, then from equations (193) through (195), the predicted intensity function is described by the GM

$$\nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}) = \sum_{j=1}^{J_{\bar{k}|\bar{k}-1}} \omega_{\bar{k}|\bar{k}-1}^{(j)} \cdot \mathcal{N}\left(\boldsymbol{x}; \mathbf{m}_{\bar{k}|\bar{k}-1}^{(j)}, \mathbf{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}^{(j)}\right). \tag{202}$$

In possession of $p_{\mathrm{C},\bar{k}|\bar{k}-1}(n)$ as in equation (192), and $\nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x})$ as in equation (202), as well as the measurement set $\boldsymbol{Z}_{\bar{k}}$, the recursion progresses with the following updating step.

Step 2 - Updating:

$$p_{\mathrm{C},\bar{k}|\bar{k}}(n) = \frac{\Upsilon_{\bar{k}}^{0}[\omega_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}](n) p_{\mathrm{C},\bar{k}|\bar{k}-1}(n)}{\langle \Upsilon_{\bar{k}}^{0}[\omega_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}], p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle}, \tag{203}$$

$$\begin{aligned}\nu_{\bar{k}|\bar{k}}(\boldsymbol{x}) = {} & \frac{\langle \Upsilon_{\bar{k}}^{1}[\omega_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}], p_{\mathrm{C},\bar{k}|\bar{k}-1}(n)\rangle}{\langle \Upsilon_{\bar{k}}^{0}[\nu_{\bar{k}|\bar{k}-1}, \boldsymbol{Z}_{\bar{k}}], p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle}[1 - p_{\mathrm{D},\bar{k}}(\boldsymbol{x})] \cdot \\ & \nu_{\bar{k}|\bar{k}-1}(\boldsymbol{x}) \sum_{\boldsymbol{z} \in \boldsymbol{Z}_{\bar{k}}} \sum_{j=1} \omega_{\bar{k}}^{(j)}(\boldsymbol{z}) \cdot \mathcal{N}\Big(\boldsymbol{x}; \mathbf{m}_{\bar{k}}^{(j)}(\boldsymbol{z}), \mathbf{\Sigma}_{\boldsymbol{x},k}^{(j)}\Big),\end{aligned} \tag{204}$$

where

$$\Upsilon_{\bar{k}}^{u}[\omega,\boldsymbol{Z}](n)=\sum_{j=0}^{\min(|\boldsymbol{Z}|,n)}(|\boldsymbol{Z}|-j)!p_{K,\bar{k}}(|\boldsymbol{Z}|-j) \tag{205}$$

$$\cdot P_{j+u}^{n}\frac{\left(1-p_{\mathrm{D},\bar{k}}(\boldsymbol{x})\right)^{n-(j+u)}}{\langle 1,\omega\rangle^{n}}e_{j}\left(\Xi_{\bar{k}}(\omega,\boldsymbol{Z})\right),$$

$$\Xi_{\bar{k}}\left(\omega,\boldsymbol{Z}\right))=\left\{\frac{\langle\mathbf{1},\kappa_{\bar{k}}\rangle}{\kappa_{\bar{k}}(\boldsymbol{z})}p_{\mathrm{D},\bar{k}}\omega^{\mathrm{T}}\mathbf{q}_{\bar{k}}(\boldsymbol{z}):\boldsymbol{z}\in\boldsymbol{Z}\right\}, \tag{206}$$

$$\omega_{\bar{k}|\bar{k}-1}=\left[\omega_{\bar{k}|\bar{k}-1}^{(1)},\ldots,\omega_{\bar{k}|\bar{k}-1}^{(J_{\bar{k}|\bar{k}-1})}\right]^{\mathrm{T}}, \tag{207}$$

$$\mathbf{q}_{\bar{k}}(\boldsymbol{z})=\left[q_{\bar{k}}^{(1)}(\boldsymbol{z}),\ldots,q_{\bar{k}}^{(J_{\bar{k}|\bar{k}-1})}(\boldsymbol{z})\right]^{\mathrm{T}}, \tag{208}$$

$$q_{\bar{k}}^{(j)}(\boldsymbol{z})=\mathcal{N}\left(\boldsymbol{z};\boldsymbol{z}_{\bar{k}|\bar{k}-1}^{(j)},\boldsymbol{\Sigma}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}\right), \tag{209}$$

$$\omega_{\bar{k}}^{(j)}(\boldsymbol{z})=p_{\mathrm{D},\bar{k}}\omega_{\bar{k}|\bar{k}-1}^{(j)}q_{\bar{k}}^{(j)}(\boldsymbol{z})\frac{\langle\Upsilon_{\bar{k}}^{1}\left[\omega_{\bar{k}|\bar{k}-1},\boldsymbol{Z}_{\bar{k}}\setminus\{\boldsymbol{z}\}\right],p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle}{\langle\Upsilon_{\bar{k}}^{0}\left[\omega_{\bar{k}|\bar{k}-1},\boldsymbol{Z}_{\bar{k}}\right],p_{\mathrm{C},\bar{k}|\bar{k}-1}\rangle}\frac{\langle\mathbf{1},\kappa_{\bar{k}}\rangle}{\kappa_{\bar{k}}(\boldsymbol{z})}, \tag{210}$$

$$\kappa_{\bar{k}}(\boldsymbol{z})=\lambda_{c}\cdot V\cdot u(\boldsymbol{z}), \tag{211}$$

$$\mathbf{m}_{\bar{k}|\bar{k}}^{(j)}(\boldsymbol{z})=\mathbf{m}_{\bar{k}|\bar{k}-1}^{(j)}+\boldsymbol{K}_{\bar{k}}^{(j)}\cdot\left(\boldsymbol{z}-\boldsymbol{\eta}_{\bar{k}|\bar{k}-1}^{(j)}\right), \tag{212}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}}^{(j)}=\mathbf{Tria}\left(\left[\mathbf{A}_{\bar{k}|\bar{k}-1}^{(j)}-\boldsymbol{K}_{\bar{k}|\bar{k}}^{(j)}\cdot\mathbf{B}_{\bar{k}|\bar{k}-1}^{(j)},\quad\boldsymbol{K}_{\bar{k}|\bar{k}}^{(j)}\cdot\boldsymbol{\Sigma}_{\boldsymbol{R},\bar{k}}^{(j)}\right]\right), \tag{213}$$

$$\boldsymbol{K}_{\bar{k}|\bar{k}}^{(j)}=\left(\mathbf{P}_{\boldsymbol{xz},\bar{k}|\bar{k}-1}^{(j)}/\left(\boldsymbol{\Sigma}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}\right)^{\mathrm{T}}\right)/\boldsymbol{\Sigma}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}, \tag{214}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}=\mathbf{Tria}\left(\left[\mathbf{B}_{\bar{k}|\bar{k}-1}^{(j)}\quad\boldsymbol{\Sigma}_{\boldsymbol{R},\bar{k}}\right]\right), \tag{215}$$

$$\mathbf{P}_{\boldsymbol{xz},\bar{k}|\bar{k}-1}^{(j)}=\mathbf{A}_{\bar{k}|\bar{k}-1}^{(j)}\cdot\left(\mathbf{B}_{\bar{k}|\bar{k}-1}^{(j)}\right)^{\mathrm{T}}, \tag{216}$$

$$\mathbf{A}_{\bar{k}|\bar{k}-1}^{(j)}=\tfrac{1}{\sqrt{N_{c}}}\left(\boldsymbol{\Psi}_{\bar{k}|\bar{k}-1}^{(j)}-\mathbf{m}_{\bar{k}|\bar{k}-1}^{(j)}\mathbf{1}_{N_{c}}^{\mathrm{T}}\right), \tag{217}$$

$$\mathbf{B}_{\bar{k}|\bar{k}-1}^{(j)}=\tfrac{1}{\sqrt{N_{c}}}\left(\boldsymbol{\Phi}_{\bar{k}|\bar{k}-1}^{(j)}-\boldsymbol{\eta}_{\bar{k}|\bar{k}-1}^{(j)}\cdot\mathbf{1}_{N_{c}}^{\mathrm{T}}\right), \tag{218}$$

$$\boldsymbol{\Psi}_{\bar{k}-1}^{(j)}=\left[\boldsymbol{\psi}_{\bar{k}-1}^{(j1)},\ldots,\boldsymbol{\psi}_{\bar{k}-1}^{(jN_{c})}\right], \tag{219}$$

$$\boldsymbol{\Phi}_{\bar{k}-1}^{(j)}=\left[\boldsymbol{\phi}_{\bar{k}-1}^{(j1)},\ldots,\boldsymbol{\phi}_{\bar{k}-1}^{(jN_{c})}\right], \tag{220}$$

$$\boldsymbol{\phi}_{\bar{k}|\bar{k}-1}^{(ji)}=g\left(\boldsymbol{\psi}_{\bar{k}|\bar{k}-1}^{(ji)}\right), \tag{221}$$

$$\boldsymbol{\psi}_{\bar{k}|\bar{k}-1}^{(ji)}=\boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}^{(j)}\boldsymbol{\varepsilon}^{(i)}+\mathbf{m}_{\bar{k}|\bar{k}-1}^{(j)}, \tag{222}$$

$$\boldsymbol{\eta}_{\bar{k}|\bar{k}-1}^{(j)}=\tfrac{1}{N_{c}}\sum_{i=1}^{N_{c}}\boldsymbol{\phi}_{\bar{k}|\bar{k}-1}^{(ji)}. \tag{223}$$

In the above, $p_{K,\bar{k}}(\cdot)$ is the distribution of the number of clutter arrivals at $\bar{k}$, $P_j^n \triangleq n!/(n-j)!$ is a permutation coefficient, $e_j(\cdot)$ is the elementary symmetric function[81], $\mathbf{\Sigma}^{(j)}_{\boldsymbol{z},\bar{k}|\bar{k}-1}$ is the square root factor of the observation covariance matrix $\mathbf{P}^{(j)}_{\boldsymbol{z},\bar{k}|\bar{k}-1}$, and $\kappa_{\bar{k}}(\cdot)$ is the clutter distribution as function of the clutter intensity λ_c, the adjusted surveillance volume V and a point process $u(\cdot)$, both of which will be specified later in Sections 6.3.2 and 6.4, respectively. Let us also point out that, since equation (213) is equivalent to the square root for the *Joseph form* of the covariance update equation [84], the Kalman gain in equation (214) is computed as two nested inverse solutions rather than the inverse of $\mathbf{P}^{(j)}_{\boldsymbol{z},\bar{k}|\bar{k}-1}$ and $\mathbf{\Sigma}^{(j)}_{\boldsymbol{R},\bar{k}}$ denotes the square root of $\boldsymbol{R}_{\bar{k}}$.

It is known that the main source of complexity in the CPHD recursion is the sum over j of elementary symmetric functions $e_j(\cdot)$ – see equation (205). Indeed it was shown in [135] that such a sum is of complexity order $\mathcal{O}\left(n \cdot |\boldsymbol{Z}_k|^3\right)$. In addition, similarly to the GM-PHD filter [118], the number of GM components propagated in time tends to grow unbounded as

$$\left(J_{\bar{k}-1} + J_{\boldsymbol{\Gamma},\bar{k}}\right) \cdot \left(1 + |\boldsymbol{Z}_{\bar{k}}|\right) = \mathcal{O}\left(J_{k|k-1} \cdot |\boldsymbol{Z}_{\bar{k}}|\right). \tag{224}$$

This indicates that a suitable strategy to reduce computational complexity of the recursion is to add, after the updating step, a *pruning* stage aimed at limiting the number of GM components by discarding those with low weights $\omega^{(j)}_{\bar{k}|\bar{k}}$ and merging the ones that are sufficiently closed together. For details of such a pruning stage we refer readers to [118].

Notice, however, that the *pruning* stage such as proposed in [118] only reduces the complexity order associated with the number of GM's, specifically by controlling $J_{\bar{k}|\bar{k}-1}$ (see equation (224)), which is minor compared to the actual complexity order $\mathcal{O}\left(n \cdot |\boldsymbol{Z}_{\bar{k}}|^3\right)$ of the filter as a whole [135]. This calls for a more effective method to reduce the complexity of the CPHD filter, which is the subject of the next section.

6.3 Gating strategies for Gaussian mixture filters

Gating is a general term used to describe techniques that *select* reduced sets of measurements $\widetilde{\boldsymbol{Z}}_{\bar{k}}$, out of the observations $\boldsymbol{Z}_{\bar{k}}$. This is typically achieved [27, pp.

[81] The elementary symmetric function can be efficiently computed using Vieta's formulas [179].

334] by defining a *validation area* on the basis of some knowledge about the noise statistics and the set of predicted measurements $\{g(\boldsymbol{X}_{\bar{k}|\bar{k}-1})\}$.

The first impact of gating is therefore to reduce the cardinality of "gated" measurement sets, since only measurements satisfying a given inequality – specifically inequality (228) in the case of Elliptical Gating, or (230) in the case of Adaptive Gating – are retained in each step. Notice that the pruning of $\boldsymbol{Z}_{\bar{k}}$ onto $\widetilde{\boldsymbol{Z}}_{\bar{k}}$ does not sacrifice the ability of the filter to handle new births, since this is accounted for during the prediction step by the augmentation of the RFS describing the predicted targets with the set $\boldsymbol{\Gamma}_{\bar{k}}$ describing the predicted births [118, 180], as from equation (59). Furthermore, taking into consideration the discussion concluding Section 6.2, it should be clear that while gating reduces the computational complexity of the filter to $\mathcal{O}\left(n \cdot |\widetilde{\boldsymbol{Z}}_{\bar{k}}|^3\right)$, it also changes the number of Gaussian Mixture components propagated in time to $\mathcal{O}\left(J_{\bar{k}|\bar{k}-1} \cdot |\widetilde{\boldsymbol{Z}}_{\bar{k}}|\right)$, in accordance with equation (224).

The second impact of gating is to affect the volume of the surveillance region V that parameterizes the clutter distribution[82] $\kappa_{\bar{k}}(\boldsymbol{z})$ in equation (211). Of course the surveillance region with gating is the *union* of all the non-empty gates, that is, all j-th gates enclosing at least one observation. If we indicate the emptiness/non-emptiness of the j-th gate by the Boolean variable $\beta^{(j)}$, then the total volume of the surveillance region with gating can be upper-bounded by the sum of validation volumes for which $\beta^{(j)} = 1$, yielding

$$V_{\mathrm{U}} = \begin{cases} \displaystyle\sum_{j=1}^{|\{\omega_{\bar{k}|\bar{k}-1}\}|} V_{\mathrm{E}}^{(j)}(n_{\boldsymbol{z}}) \cdot \beta^{(j)} & \text{for Elliptical Gating,} \qquad (225a) \\ \displaystyle\sum_{j=1}^{|\{\omega_{\bar{k}|\bar{k}-1}\}|} V_{\mathrm{A}}^{(j)}(n_{\boldsymbol{z}}) \cdot \beta^{(j)} & \text{for Adaptive Gating,} \qquad (225b) \end{cases}$$

where the quantities $V_{\mathrm{E}}^{(j)}(n_{\boldsymbol{z}})$ and $V_{\mathrm{A}}^{(j)}(n_{\boldsymbol{z}})$ will be defined in the following subsections.

Notice that as a consequence of upper-bounding, V_{U} can be larger than the original surveillance region, which we shall denote by V_{S}. With that in mind, for consistency, the effective surveillance region V to be used in equation (211) can

[82] In the simulations shown in Section 6.4 the clutter distribution $\kappa_{\bar{k}}(\boldsymbol{z})$ is modeled by a Poisson RFS with intensity λ_c.

be finally determined by

$$V = \min\{V_{\mathrm{U}}, V_{\mathrm{S}}\}. \tag{226}$$

By associating with each predicted point a "feasible region" (the gate), and discarding measurements incompatible with such a region/gate the impact of spurious measurements is reduced and the possible estimation outcomes are restricted. Therefore the effect of using gates of sufficiently large sizes is, with a high probability, to reduce clutter without causing any further impact on the filter's performance.

6.3.1 Elliptical gating

Refer to the measurement model described in equation (176b) and let $\boldsymbol{\epsilon}_{\bar{k}}^{(ij)}$ denote the *innovation* vector associated with the i-th measurement $\boldsymbol{z}_{\bar{k}}^{(j)} \in \boldsymbol{Z}_{\bar{k}}$ and with respect the j-th predicted observation, *i.e.*

$$\boldsymbol{\epsilon}_{\bar{k}}^{(ij)} \triangleq \boldsymbol{z}_{\bar{k}}^{(i)} - g\left(\boldsymbol{x}_{\bar{k}|\bar{k}-1}^{(j)}\right). \tag{227}$$

Under the assumption that $\boldsymbol{r}_{\bar{k}}$ in equation (176b) is an additive zero mean white Gaussian noise, the residual covariance matrix of $\boldsymbol{\epsilon}_{\bar{k}}^{(ij)}$ is known to be[83] $\mathbf{P}_{\boldsymbol{\epsilon},\bar{k}|\bar{k}-1}^{(j)} = \boldsymbol{G} \cdot \mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1}^{(j)} \cdot \boldsymbol{G}^{\mathrm{T}} + \boldsymbol{R}$, where $\mathbf{P}_{\boldsymbol{x},\bar{k}|\bar{k}-1}^{(j)} = \boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}^{(j)} \cdot \left(\boldsymbol{\Sigma}_{\boldsymbol{x},\bar{k}|\bar{k}-1}^{(j)}\right)^{\mathrm{T}}$ is the predicted covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{R}$ the measurement noise covariance matrix. Notice also that with respect to the S-CKF implementation of the GM-CPHD offered in Section 6.2, $\mathbf{P}_{\boldsymbol{\epsilon},\bar{k}|\bar{k}-1}^{(j)}$ is the covariance matrix $\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}$. Therefore, for the sake of clarity in the following the innovation covariance matrix will be still referred to as $\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}$.

It follows that the Mahalanobis distance associated with $\boldsymbol{\epsilon}_{\bar{k}}^{(ij)^{\mathrm{T}}}$ is $d_g^2 \triangleq \boldsymbol{\epsilon}_{\bar{k}}^{(ij)^{\mathrm{T}}} \cdot \left(\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}\right)^{-1} \cdot \boldsymbol{\epsilon}_{\bar{k}}^{(ij)}$. In the case of linear Gaussian systems, ellipsoidal validation gates (a.k.a elliptical gates) are known to be optimal [181] and can be easily computed. Specifically, if T_g denotes the $n_{\boldsymbol{z}}$-dimensional validation threshold, then elliptical gating amounts to selecting the measurements $\boldsymbol{z}_{\bar{k}}^{(i)}$ as

[83] When $g(\boldsymbol{x})$ is a nonlinear function of $\boldsymbol{x}$, $\boldsymbol{G}$ can be recovered with a simple first order approximation, that is $\boldsymbol{G} = \partial g/\partial \boldsymbol{x}$ calculated at $\boldsymbol{x} = \boldsymbol{x}_{\bar{k}|\bar{k}-1}$ or by means of the cubature transform [84].

follows:

$$\boldsymbol{z}_{\bar{k}}^{(i)} \begin{cases} \in \widetilde{\boldsymbol{Z}}_{\bar{k}}^{(i)} & \text{if } \exists\,(i,j) \mid \boldsymbol{\epsilon}_{\bar{k}}^{(ij)^{\mathrm{T}}} \cdot \left(\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}\right)^{-1} \cdot \boldsymbol{\epsilon}_{\bar{k}}^{(ij)} \leqslant T_g, \\ \notin \widetilde{\boldsymbol{Z}}_{\bar{k}}^{(i)} & \text{otherwise.} \end{cases} \tag{228}$$

Therefore, under the assumption that both noise processes $\boldsymbol{q}$ and $\boldsymbol{r}$ are Gaussian, as well as the innovation vector $\boldsymbol{\epsilon}$, the optimal value for the threshold T_g is the α quantile of the upper-tail of a chi-square distribution with $n_{\boldsymbol{z}}$ degrees of freedom [27, pp. 337]. In summary, due to the Gaussian assumption over $\boldsymbol{\epsilon}^{(ij)}$, $d_g^2 \sim \chi_{n_{\boldsymbol{z}}}^2$, where $\chi_{n_{\boldsymbol{z}}}^2$ denotes a χ^2 distribution with $n_{\boldsymbol{z}}$ degrees of freedom, and T_g is a threshold recovered from χ^2 distribution tables. Notice also that this is equivalent to setting the probability that a correct observation falls within the gate T_g to be $P_g = 1 - \alpha$.

Let $c_{n_{\boldsymbol{z}}}$[84] denote the surface of the unit hypersphere in the $n_{\boldsymbol{z}}$ dimensional space. Then, the volume for the ellipsoidal gate is

$$V_{\mathrm{E}}^{(j)}(n_{\boldsymbol{z}}) \triangleq c_{n_{\boldsymbol{z}}} \cdot \left[\det(\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}) \cdot T_g^{n_{\boldsymbol{z}}}\right]^{\frac{1}{2}}. \tag{229}$$

where we have omitted from $V_{\mathrm{E}}(n_{\boldsymbol{z}})$ the sub index $_{\bar{k}|\bar{k}-1}$ for the sake of notational simplicity.

It is clear from equation (228) that gating is performed for each i-th measurement and against each j-th GM component. Notice that in equation (229) the determinant of the residual covariance matrix $\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}$ needs to be computed repeatedly. Under the square-root version of the CKF propagation of the GM-CPHD filter described in Section 6.2, instead of propagating the covariances $\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}$ themselves, it is their triangular components $\boldsymbol{\Sigma}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}$ that are propagated, which enables the fast evaluation of such determinants. In other words, some complexity reduction on elliptical gating itself is also achieved as a by-product of the contribution of Section 6.2.

6.3.2 Adaptive gating for GM filters

Although not emphasized before, one problem with the gating method discussed above is to find the right value for T_g, or equivalently, through the α quantile of

[84] As shown in [27, pp. 338], $c_{n_{\boldsymbol{z}}} = \frac{\pi^{n_{\boldsymbol{z}}/2}}{\Gamma(\frac{n_{\boldsymbol{z}}}{2}+1)}$, which for the particular cases of $n_z = 1$ and $n_{\boldsymbol{z}} = 2$ yield $c_{n_{\boldsymbol{z}}} = 2$ and $c_{n_{\boldsymbol{z}}} = \pi$, respectively.

the upper-tail of a χ^2 distribution with $n_{\boldsymbol{z}}$ degrees of freedom [27, pp. 337], a proper value for P_g. Indeed when T_g (namely P_g) is set too large, many false alarm measurements will be selected (not gated), with clear impact on the computational load of the algorithms. On the other hand, if the gate size is too small, some of the measurements actually generated by targets might be removed by the gate with potentially severe consequences on performance.

The aforementioned problem is the quintessential problem of gating techniques, namely, they can reduce complexity but are sensitive to parameterization. In addition to elliptical gating, other previously proposed gating methods also exhibit the same drawback [181, 182]. In this section we propose an alternative gating method that is robust to parameterization. The key to our gating technique is to *adaptively* change the gate associated with each predicted measurement $g(\boldsymbol{x}_{\bar{k}|\bar{k}-1}^{(j)})$ by increasing the corresponding innovation covariance matrix $\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}$ or equivalently augmenting the threshold of the corresponding Mahalanobis distance by the $\omega_{\bar{k}|\bar{k}-1}^{(j)}$ used in the GM implementation of the filters. Mathematically, the adaptive gating scheme is summarized as follows:

$$\boldsymbol{z}_{\bar{k}}^{(i)} \begin{cases} \in \widetilde{\boldsymbol{Z}}_{\bar{k}}^{(i)} & \text{if } \exists\,(i,j) \mid \boldsymbol{\epsilon}_{\bar{k}}^{(ij)^{\mathrm{T}}} \cdot \left(\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}\right)^{-1} \cdot \boldsymbol{\epsilon}_{\bar{k}}^{(ij)} \leqslant T_g \cdot \left(1+\omega_{\bar{k}|\bar{k}-1}^{(j)}\right), \\ \notin \widetilde{\boldsymbol{Z}}_{\bar{k}}^{(i)} & \text{otherwise.} \end{cases} \tag{230}$$

As implied by inequality (230), in the adaptive gating method the gate size is adjusted not only with basis on the noise statistics (captured by T_g), but also depending on the likelihood of the predicted points, as measured by the GM weights $\omega_{\bar{k}|\bar{k}-1}^{(j)}$.

Rewriting equation (229) in light of inequality (230), the volume of the j-th adaptive gate becomes

$$V_{\mathrm{A}}^{(j)}(n_{\boldsymbol{z}}) = c_{n_{\boldsymbol{z}}} \cdot \left[\det(\mathbf{P}_{\boldsymbol{z},\bar{k}|\bar{k}-1}^{(j)}) \cdot \left(T_g \cdot (1+\omega_{\bar{k}|\bar{k}-1}^{(j)})\right)^{n_{\boldsymbol{z}}}\right]^{\frac{1}{2}}. \tag{231}$$

While it is true that if compared individually the adaptive gates are always larger than the elliptical one, elliptical gates are all of the same size, so that the separation of spurious data from real measurements is not carried out in a way that discriminates the variable likelihood these have. In contrast, in the proposed technique, gates are larger around points with higher likelihood, and smaller around points with lower, such that the overall effect is that it is possible to be more aggressive in regions where measurements are more likely to be

clutter, and more conservative where measurements are more likely to be truly generated from targets.

6.4 Performance evaluation and comparisons

In the following the performance achieved by the gating techniques discussed in Section 6.3 applied to the S-CKF-GM-CPHD filter are compared amongst each other. To do so the two different scenarios represented in Figures 29-30 are considered.

Specifically, the first scenario consists of a multisensor multitarget setting in which the generic i-th target at time $\bar{k}$ is described by the state vector $\boldsymbol{x}_{\bar{k}}^{(i)} = [\rho_{\bar{k}}^{(i)}, \varrho_{\bar{k}}^{(i)}, \dot{\rho}_{\bar{k}}^{(i)}, \dot{\varrho}_{\bar{k}}^{(i)}]^{\mathrm{T}}$ with $\rho_{\bar{k}}^{(i)}$ and $\varrho_{\bar{k}}^{(i)}$ denoting positions in the $x-y$ plane and $\dot{\rho}_{\bar{k}}^{(i)}$ and $\dot{\varrho}_{\bar{k}}^{(i)}$ the corresponding velocity coordinates. Trajectories are assumed to evolve accordingly to a nearly constant velocity (NCV) [30] with transition and process noise matrix

$$\mathbf{F}_{\bar{k}} = \begin{bmatrix} \mathbf{I}_2 & \bar{\Delta} \cdot \mathbf{I}_2 \\ \mathbf{0}_{2\times 2} & \mathbf{I}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\boldsymbol{Q}} = \sigma_q \begin{bmatrix} \frac{\bar{\Delta}^2}{2} \cdot \mathbf{I}_2 \\ \bar{\Delta}\mathbf{I}_2 \end{bmatrix}, \tag{232}$$

where $\mathbf{0}_{2\times 2}$ is a 2×2 matrix of zeros, $\bar{\Delta}$ is the sampling time and $\sigma_q = 12$ m/s^2 represents the standard deviation of the process noise.

The target birth process is assumed to be Poisson distributed with $\nu_{\Gamma,\bar{k}} = \omega_\Gamma \mathcal{N}(\boldsymbol{x}; \mathbf{m}_1, \sigma_\Gamma \cdot \mathbf{I}_4) + \omega_\Gamma \mathcal{N}(\boldsymbol{x}; \mathbf{m}_2, \sigma_\Gamma \cdot \mathbf{I}_4)$, where $\omega_\Gamma = 0.1$, $\mathbf{m}_1 = [250, 250, 0, 0]$, $\mathbf{m}_2 = [-250, -250, 0, 0]$ and $\sigma_\Gamma = 10$. Targets trajectories are shown in Figure 29, from where it can be understood that at $\bar{k} = 1$ two targets with 100 steps long trajectories generate from the points $[250, 250]$ and $[-250, -250]$ respectively and at $\bar{k} = 47$ a new target appears from the point $[250, 250]$.

Range measurements are performed at fixed devices (or sensors) whose absolute location is assumed to be known and that in the following will be therefore referred to as anchor nodes. The range observation at the a-th anchor and with respect to the i-th target is

$$z_{a,k}^{(i)} \triangleq \sqrt{\left(\rho_k^{(i)} - \rho^{(a)}\right)^2 + \left(\varrho_k^{(i)} - \varrho^{(a)}\right)^2} + r_k, \tag{233}$$

where $r_{\bar{k}} \sim \mathcal{N}(\cdot; 0, \sigma_r)$ is the measurement noise and where $\sigma_r = 10$.

Clutter measurements are generated accordingly to a Poisson process with $\lambda_c = 4 \times 10^{-3}$ over a measurement space of $2700m$, resulting in an average of 10.8 false detection at each anchor per time $\bar{k}$.

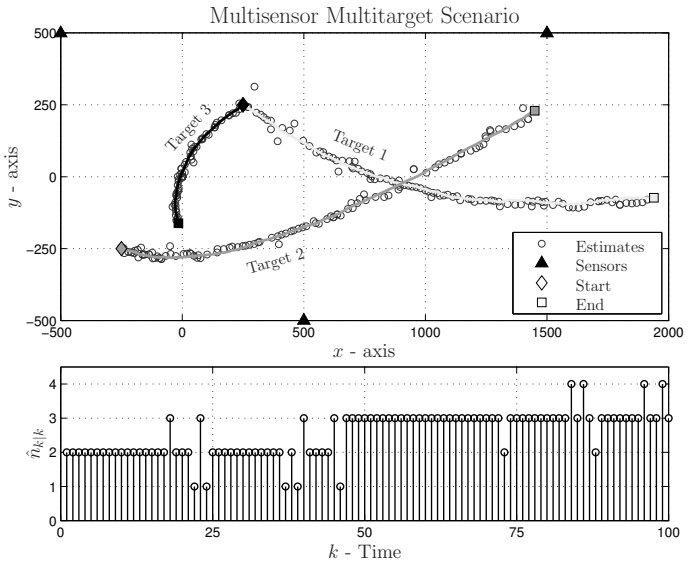


Fig 29. Multitarget multisensor scenario described in Section 6.4. Each anchor is subject to an average of 10 clutter measurements per time and the third targets appears at $\bar{k} = 47$. Revised form [20].

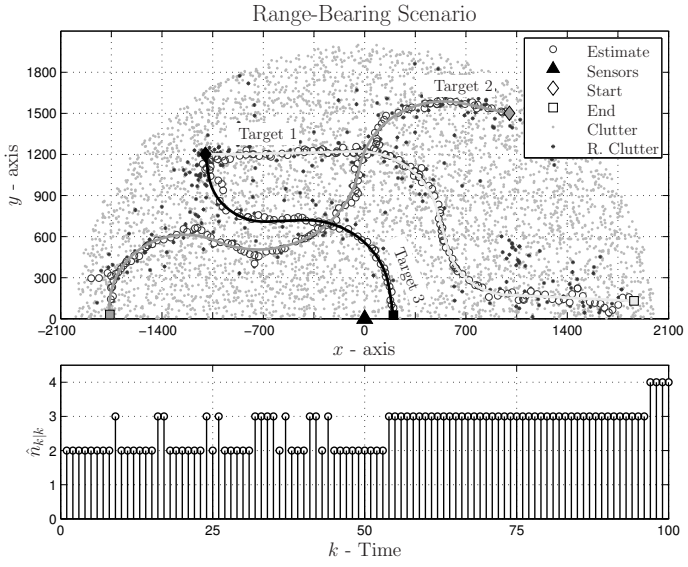
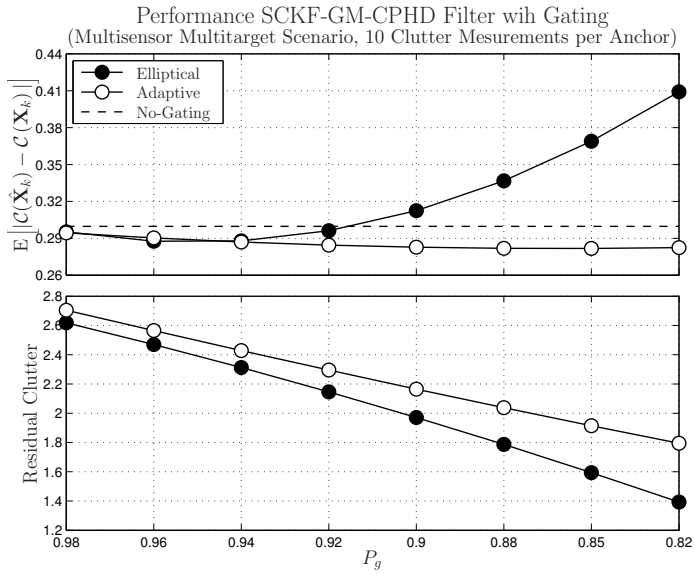
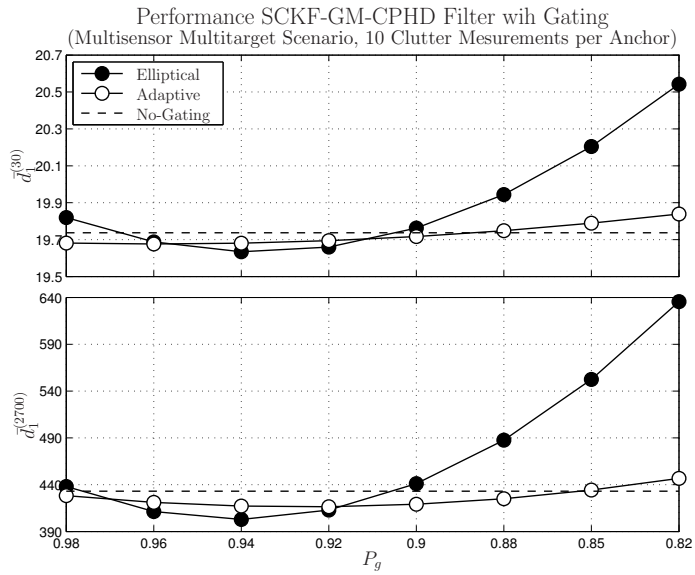


Fig 30. Sample scenario where the first two targets cross at time $\bar{k} = 33$ and the third target appears at $\bar{k} = 51$. The average number of clutter measurements per time is set to 50. Revised form [13].



(a) Cardinality error and residual clutter for the CKF-GM-CPHD filter.



(b) Average OSPA metric for the CKF-GM-CPHD filter.

Fig 31. Results for the CKF-GM-CPHD applied to the scenario in Figure 29 with an average of 10 clutter measurements per time and $\boldsymbol{r}_{\bar{k}} \sim \mathcal{N}(x; 0, \sigma_r)$ with $\sigma_r = 10$.

The probability of target detection and survival are set to $p_{\mathrm{D}} = 0.98$ and $p_{\mathrm{S}} = 0.99$ respectively. Following the performance for the S-CKF-GM-CPHD filter used together with the gating strategies discussed in Section 6.3 are studied as a function of P_g.[85]

Specifically, Figure 31(a) compares the error in the estimated cardinality as well the residual clutter[86] while Figure 31(b) shows the average OSPA metric over the entire tracking period for the cutoff values $c = 30$, chosen to capture the *localization* errors; and $c = 2700$ chosen to describe the impact of *cardinality* errors.

The results show that by gating it is possible to remove most of the clutter measurements in $\boldsymbol{Z}_{\bar{k}}$ without compromising the filter performance. In particular, elliptical gating causes performance to rapidly deteriorate for $P_g < 0.92$, whereas with adaptive gating performance is not affected. Moreover, as shown later, whenever elliptical gating is fed with erroneous estimates of the measurement error covariance matrix $\boldsymbol{R}$, it is more likely to discard target generated measurements, resulting in a severe degradation of the filter performance.

The second scenario consists of a single sensor multitarget setting in which targets are assumed to evolve according to the following nearly coordinated turning model [30, pp. 467]

$$\boldsymbol{x}_k = \mathbf{F}(\Omega) \cdot \boldsymbol{x}_{k-1} + \boldsymbol{\Sigma}_{\boldsymbol{Q}}, \tag{234}$$

where the state vector $\boldsymbol{x}_{\bar{k}}$ is modified to include the turning rate Ω as fifth state and where the transition and process noise matrices are

$$\mathbf{F}(\Omega) = \begin{bmatrix} 1 & 0 & \frac{\sin(\Omega \cdot \bar{\Delta})}{\Omega} & -\frac{1-\cos(\Omega \cdot \bar{\Delta})}{\Omega} & 0 \\ 0 & 1 & \frac{1-\cos(\Omega \cdot \bar{\Delta})}{\Omega} & \frac{\sin(\Omega \cdot \bar{\Delta})}{\Omega} & 0 \\ 0 & 0 & \cos(\Omega \cdot \bar{\Delta}) & -\sin(\Omega \cdot \bar{\Delta}) & 0 \\ 0 & 0 & \sin(\Omega \cdot \bar{\Delta}) & \cos(\Omega \cdot \bar{\Delta}) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{235}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{Q}} = \begin{bmatrix} \frac{\bar{\Delta}^2}{2} \cdot \mathbf{I}_2 & \mathbf{0}_2 \\ \bar{\Delta} \cdot \mathbf{I} & \mathbf{0}_2 \\ \mathbf{0}_2^{\mathrm{T}} & \bar{\Delta} \end{bmatrix} \cdot \begin{bmatrix} \sigma_q \cdot \mathbf{1}_2 \\ \sigma_\Omega \end{bmatrix}, \tag{236}$$

[85] Since $z_{a,\bar{k}}^{(i)} \in \mathbb{R}^{n_{\boldsymbol{z}}}$ where $n_{\boldsymbol{z}} = 1$, then P_g and T_g are related by: $P_g \triangleq \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{T_g}} \exp(-u^2/2) \cdot du$.

[86] The term "residual clutter" refers to the difference between the average clutter intensity $\lambda_c \cdot V$ and the average number of observations discarded from $\boldsymbol{Z}_{\bar{k}}$ by gating.

where $\sigma_q = 12$ m/s^2 and $\sigma_\Omega = 2\pi/180$ rad/s are related to the process noise intensity and $\mathbf{0}_2$ is a column vector of zeros of length 2.

Assuming the measuring sensor at the origin of the reference system, then at time $\bar{k}$ the observation generated by the i-th target consists of a range-bearing measurement

$$\boldsymbol{z}_{\bar{k}} = \begin{bmatrix} \sqrt{\left(\rho_{\bar{k}}^{(i)}\right)^2 + \left(\varrho_{\bar{k}}^{(i)}\right)^2} \\ \tan^{-1}\left(\dfrac{\rho_{\bar{k}}^{(i)}}{\varrho_{\bar{k}}^{(i)}}\right) \end{bmatrix} + \boldsymbol{r}_{\bar{k}}, \tag{237}$$

where $\boldsymbol{r}_{\bar{k}} \sim \mathcal{N}(\mathbf{r}; 0, \boldsymbol{\Sigma}_{\boldsymbol{R}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{R},\bar{k}} = \operatorname{diag}([\sigma_r, \sigma_\theta])$, $\sigma_\theta = 2(\pi/180)$ and $\sigma_r = 10$.

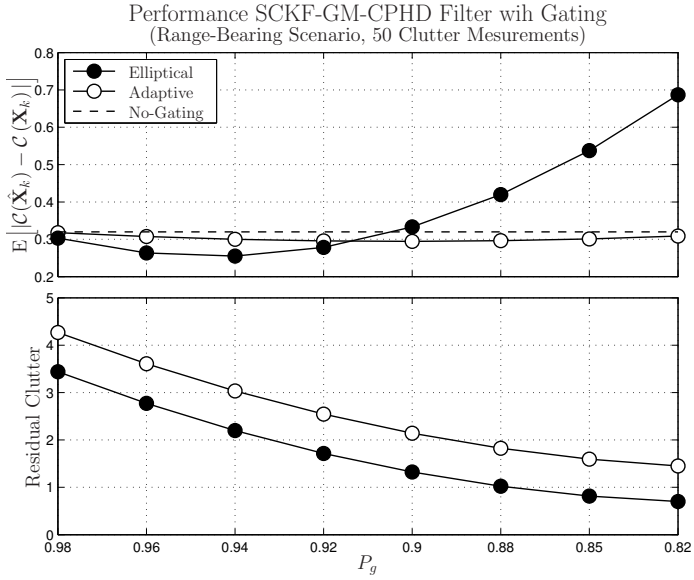
The birth process is assumed to be a Poisson and it is modeled by the GM $\nu_{\Gamma,k}(\boldsymbol{x}) = \omega_\Gamma \mathcal{N}(\boldsymbol{x}; \boldsymbol{m}_\Gamma^{(1)}, \boldsymbol{\Sigma}_\Gamma) + \omega_\Gamma \mathcal{N}(\boldsymbol{x}; \boldsymbol{m}_\Gamma^{(2)}, \boldsymbol{\Sigma}_\Gamma)$ with $\omega_\Gamma = 0.1$, $\mathbf{m}_\Gamma^{(1)} = [-1100, 1200, 0, 0, 0]^{\mathrm{T}}$, $\mathbf{m}_\Gamma^{(2)} = [1000, 1500, 0, 0, 0]^{\mathrm{T}}$ and the square-root of the birth-covariance matrix given by $\boldsymbol{\Sigma}_\Gamma = \operatorname{diag}([10, 10, 7, 7, 7(\pi/180)])$.

Both p_{D} and p_{S} have the same values than in the multisensor scenario described above while the clutter intensity is set to $\lambda_c = 8 \times 10^{-3}$ over a surveillance region of $[-\pi/2, \pi/2]$ rad $\times$ $[0, 2000]$ m, resulting in an average of 50 clutter measurements per time. Figure 30 shows the targeted scenario, where at $\bar{k} = 1$ two targets having 100 steps long trajectories are generated from the points $[-1100, 1200]$ and $[1000, 1500]$ and cross at $\bar{k} = 33$. A new target appears at $\bar{k} = 51$ from the point $[-1100, 1200]$. Similarly to the multisensor scenario, the performance for the gating strategies discussed in Section 6.3 and applied to the S-CKF-GM-CPHD filter are compared for different values of P_g.[87]

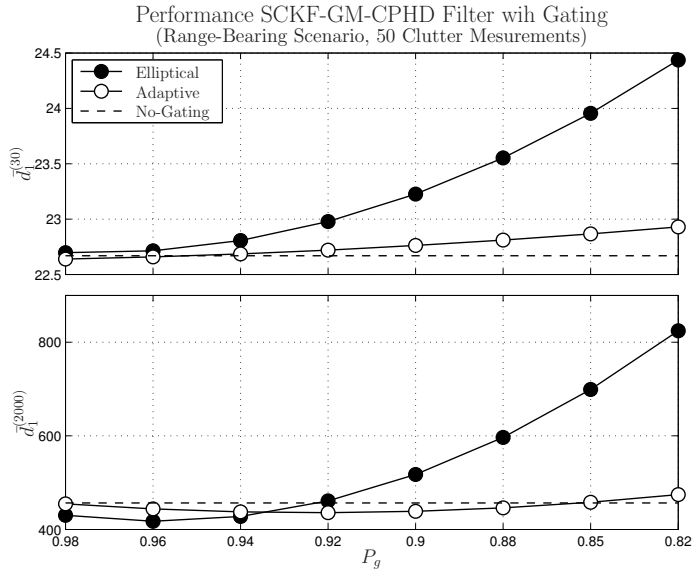
Figure 32(a) compares the error in the estimated cardinality as well the residual clutter while Figure 32(b) the average OSPA metric over the entire tracking period computed for the two cutoff values $c = 30$ and $c = 2000$.

Once again by gating it is possible to remove most of the clutter measurements in $\boldsymbol{Z}_{\bar{k}}$ without compromising the filter performance and more importantly elliptical gating causes performance to rapidly deteriorate for $P_g < 0.92$, whereas the adaptive strategy does not. Moreover, as Figure 33 shows, whenever the gating strategies are fed with erroneous estimates of the measurement error covariance matrix $\boldsymbol{R}$, the elliptical gate always deteriorates the tracking performance.

[87] Since $z_{a,\bar{k}}^{(i)} \in \mathbb{R}^{n_{\boldsymbol{z}}}$ where $n_{\boldsymbol{z}} = 2$, then P_g and T_g are related by: $P_g \triangleq 1 - \exp(-T_g/2)$.



(a) Cardinality error and residual clutter for the CKF-GM-CPHD filter.



(b) Average OSPA metric for the CKF-GM-CPHD filter.

Fig 32. Results for the S-CKF-GM-CPHD applied to the range-bearing scenario described in Section 6.4 and represented in Figure 30. Revised form [13].

Differently adaptive gating is still able to reach the performance of the achieved by the S-CKF-GM-CPHD filter at a considerably lower computational time. A time comparison between for the S-CKF-GM-CPHD filter with and without the adaptive gating is offered in Table 2.

Table 2. Computational time[a].

	Multisensor	Range-bearing
No Gating	$\tau = 150.65$	$\tau = 198.5$
Adaptive		
$P_g = 0.98$	$\tau = 96.0$	$\tau = 29.50$
$P_g = 0.94$	$\tau = 91.5$	$\tau = 24.60$
$P_g = 0.90$	$\tau = 85.5$	$\tau = 20.95$
$P_g = 0.85$	$\tau = 77.1$	$\tau = 18.26$

[A] Time τ in seconds and for the same machine.

[B] For the Scenarios described in Section 6.4.

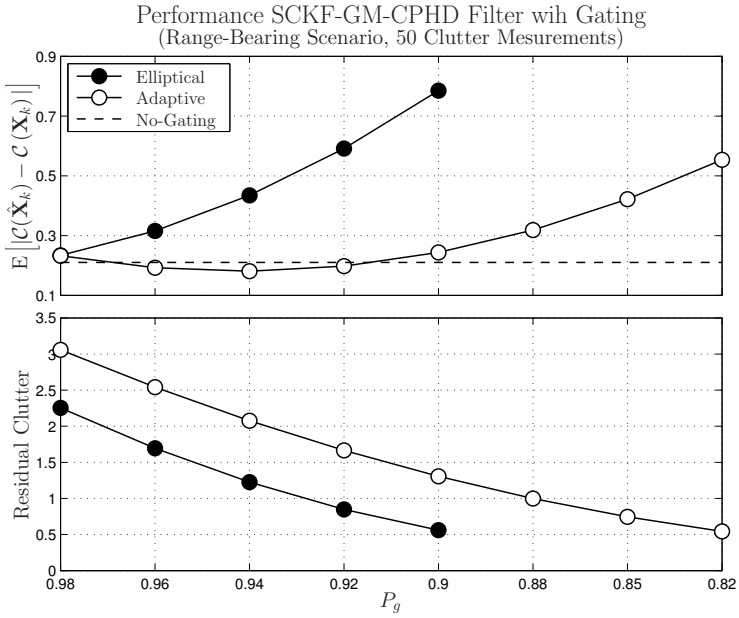


Fig 33. Results for the S-CKF-GM-CPHD applied to the range-bearing scenario described in Section 6.4 and represented in Figure 30 where the value of σ_r in $\boldsymbol{R}$ used in the filter is 7 rather than the true standard deviation on the range error 10.

6.5 Summary and discussions

The chapter showed that under the framework here considered – namely the Gaussian Mixture CPHD, in which Gaussianity is ensured by construction – the integration of the square-root CKF into the GM-CPHD recursion is not only more suitable (*i.e.* the best design under the Gaussian assumption), to state of the art solutions, *e.g.* the EKF.

Moreover, by *adaptively* changing the gate associated with each predicted measurement $g(\boldsymbol{x}^{(j)}_{\bar{k}|\bar{k}-1})$ in proportion to the likelihood of the single GM components used in the filters, it is possible to achieve the same run time improvement than elliptical gates without degrading the results.

It must be emphasized that the proposed gating strategy is applicable (essentially) to **any** GM filter, including (but not exclusively) the GM-(C)PHD filters and their variations, such as the GM implementation of the CB-MeMBer filter proposed in [183]. In conclusion the proposed adaptive gating applied to the S-CKF-GM-CPHD filter offers better robustness, numerical stability and, as shown in Table 2 lower complexity than the non-gated GM-CPHD filter.

7 Conclusions and future works

7.1 Summary and conclusions

Due to diversity of the scenarios requiring positioning applications, a unique solution to the multitarget LT problem is not realistic. In this work several state of the art algorithms to solve the MTT problem have been improved upon to provide a lower computational complexity, better accuracy or wider range of applicability.

In particular, in the context of *active* multitarget scenarios a comprehensive algorithm to handle the problem of simultaneously tracking a large number of targets with no *a priori* mobility models was proposed. This was accomplished by modifying the Jacobi algorithm used to eigen-decompose the subsequent Euclidean kernels in the C-MDS algorithm which in turn was based on the statistical characterization of the spectrum of the random Gramian matrices used in the algorithm, achieved by employing an extreme-value and asymptotic take on the theory of Gershgorin spectrum bounds.

The results revealed that, even for relatively large subspace distance values, the matrix can, with a high probability (quantified analytically), be eigen-decomposed by employing a statistically optimized number of plane rotations. Moreover, we showed that the Jacobi algorithm used together with the proposed truncated LM sweeping strategy achieves the convergence rate of the optimal strategy applied to but without requiring the search for the optimal element to be annihilated.

To extend the application scenario of the MDS-based method to incomplete Euclidean distance matrices, an efficient Nyström-inspired method to compensate for data erasures was proposed in Chapter 5 where the problem of heterogeneous information was also addressed by extending the C-MDS framework to work on the edges of the graph generated by the target nodes.

To improve the computational time required by state-of-the-art solutions to the *passive* MTT problem in Chapter 6 it was proposed to integrate the S-CKF into the GM-CPHD recursion which is numerically superior to state of the art solutions, *e.g.* the EKF and UKF.

Moreover, by adaptively changing the gate associated with the predicted measurements in proportion to the likelihood of the single GM components used in the filters, it is possible to achieve the same run time improvement as elliptical gating but without degrading the results. Although tested in range-only multitarget multisensor and a baring single sensor MTT scenario, the approach is applicable to any GM filter, including (but not exclusively) the GM-(C)PHD filters and their variations.

In conclusion the proposed adaptive gating applied to the S-CKF-GM-CPHD filter offers better robustness, numerical stability and lower complexity than the standard non-gated GM-CPHD filter.

7.2 Future works

Although the algorithms proposed in Chapters 4 and 5 represent a comprehensive solution to handle the problem of simultaneously tracking a large number of targets with no *a priori* mobility models, similarly to the state covariance matrix in the KF algorithm, some application scenarios might require feedback on the quality of the location estimates.

One way to include such information as well as eventual knowledge about the targets' dynamic, is to add to the proposed C-MDS-based solutions an additional stage based on the *interval analysis* tools. Indeed, in [172] it was already shown that in tracking applications this framework allows to propagate in time, instead of punctual estimates, sets of bounds which are proportional to the uncertainty of the solution. Another point of future investigation, is the generalization of the dissimilarity measures used in the subspace MDS-based algorithms discussed in Chapter 4 and 5 to include, instead of simple range and angle information, different kind of similarities, *e.g.* connectivity and or semantic information. In this way it would be possible to extract in an efficient way information not necessarily related to positions but anyway important in the more general case of context-aware scenarios.

Differently, the analysis provided in Chapter 4 could be generalized to the case of complex squared matrices, *i.e.* Hermitian matrices which could be useful in a wide range of applications in communication (*e.g.* [184]).

Concerning the *passive* LT scenarios, the adaptive gating strategy proposed in Chapter 6 could be applied, together with the S-CKF-GM-CPHD filter, to

the recently proposed Cardinalized Balanced Multi-Target Multi-Bernoulli (CB-MeMBer) filter [183]. In addition, since the proposed gating strategy relaxes the need for precise clutter rate information, in the future it is intended to compare the proposed gated CPHD solution against algorithms that address this specific problem directly, *e.g.* [185].

More importantly, due to the nature of the problem, namely a continuous process sampled at discrete times, an additional point of interest for future research activities would be to consider, instead of the *discrete-discrete* solution presented in Chapter 6, a *continuos-discrete* filter, such as the ones discussed in [134] and study the eventual gain in performance versus the increase in complexity.

References

1. Yu F, Wong VWS & Leung VCM (2005) Performance enhancement of combining QoS provisioning and location management in wireless cellular networks. IEEE Transactions on Wireless Communications 4(3): 943–953.
2. Bshara M, Orguner U, Gustafsson F & Van Biesen L (2010) Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on WIMAX networks. IEEE Transaction on Vehicular Technology 59(1): 283–294.
3. Kant K, Udar N & Viswanathan R (2008) Enabling location-based services in data centers. IEEE Network 22(6): 20–25.
4. Mazuelas S, Laga F, Blas J, Bahillo A, Fernandez P, Lorenzo RM & Abril EJ (2009) Prior NLOS measurement correction for positioning in cellular wireless networks. IEEE Transaction on Vehicular Technology 58(5): 2585–2951.
5. Yan C, Fan HH & Martin RK (2009) Robust NLoS mitigation for TOA estimation. In: IEEE International Conference on Acoustics, Speech and Signal Processing: 2525–2528.
6. Heidari M, Alsindi NA & Pahlavan K (2009) UDP identification and error mitigation in TOA-based indoor localization systems using neural network architecture. IEEE Transaction on Wireless Communications 8(7): 3597–3607.
7. Dardari D, Conti A, Ferner U, Giorgetti A & Win MZ (2009) Ranging with ultrawide bandwidth signals in multipath environments. IEEE Proceedings 97(2): 404–426.
8. Marano S, Gifford W, Wymeersch H & Win M (2010) NLOS identification and mitigation for localization based on UWB experimental data. IEEE Journal on Selected Areas in Communications 28(7): 1026–1035.
9. Chen CH, Feng KT, Chen CL & Tseng PH (2009) Wireless location estimation with the assistance of virtual base stations. IEEE Transaction on Vehicular Technology 58(1): 93–106.
10. Destino G & de Abreu GTF (2009) Weighing strategy for network localization under scarce ranging information. IEEE Transaction on Wireless Communications 8(7): 3668–3678.
11. Mazuelas S, Lorenzo R, Bahillo A, Fernandez P, Prieto J & Abril E (2010) Topology assessment provided by weighted barycentric parameters in harsh environment wireless location systems. IEEE Transactions on Signal Processing 58(7): 3842–3857.
12. Macagnano D & de Abreu GTF (2011) Gershgorin analysis of random Gramian matrices with application to MDS tracking. IEEE Transactions on Signal Processing 59(4): 1785–1800.
13. Macagnano D & de Abreu GTF (2011) Adaptive gating for multitarget tracking with Gaussian mixture filters. IEEE Transactions on Signal Processing 60(3): 1533–1538.

14. Macagnano D & de Abreu GTF (2009) A wavelet pre-filter scheme for MDS tracking. In: IEEE Workshop on Positioning, Navigation and Communication: 301–306.
15. Macagnano D & de Abreu GTF (2009) Weighting strategy for MDS tracking in los conditions. In: IEEE Personal Indoor Mobile Radio Communication (PIMRC): 2833–2837.
16. Macagnano D & de Abreu GTF (2010) Gershgorin discs of gramian matrices for target tracking. In: IEEE Workshop on Positioning, Navigation and Communication: 50–55.
17. Macagnano D & de Abreu GTF (2010) Statistical spectral analysis of random Gramian matrices. In: IEEE Asilomar Conference on Signals, Systems, and Computers: 1802–1806.
18. Macagnano D & de Abreu GTF (2011) Super MDS with heterogeneous information. In: IEEE Asilomar Conference on Signals, Systems, and Computers.
19. Macagnano D, Destino G & de Abreu GTF (2011) Hybrid initialization for non-convex network localization problems. In: IEEE International Conference on Ultra-Wideband: 145–149.
20. Macagnano D & de Abreu GTF (2010) Multitarget tracking with the cubature Kalman probability hypothesis density filter. In: IEEE Asilomar Conference on Signals, Systems and Computers: 1455–1459. Pacific Grove, CA, U.S.A.
21. Macagnano D & de Abreu GTF (2011) Gating for multitarget tracking with the Gaussian mixture PHD and CPHD filters. In: IEEE Workshop on Positioning, Navigation and Communication: 149–154.
22. Macagnano D & de Abreu GTF (2011) On the spectrum of perturbed Gramian matrices. In: IEEE Communication Theory Workshop.
23. de Abreu GTF & Destino G (2007) Super MDS: Source location from distance and angle information. In: Wireless Communications and Networking Conference: 4430–4434.
24. Gu Y, Lo A & Niemergeers I (2009) A survey of indoor positioning systems for wireless personal networks. IEEE Commununications Surveys & Tutorials 11(1): 13–32.
25. Liu H, Darabi H, Banerjee P & Liu J (2007) Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews 37(6): 1067–1080.
26. Priyantha NB (2005) The cricket indoor location system. Ph.D. thesis, Massachusetts Institute of Technology.
27. Blackman S & Popoli R (1999) Design and Analysis of Modern Tracking Systems. Artech House.
28. Ash JN & Moses RL (2009) Self-localization of Sensor Networks. in Handbook on Array Processing and Sensor Networks, IEEE-Wiley.
29. Macagnano D, Destino G, Esposito F & Abreu G (2007) MAC performances for localization and tracking in wireless sensor networks. In: IEEE Workshop on Positioning, Navigation and Communication: 297–302.
30. Bar-Shalom Y, Li XR & Kirubarajan T (2001) Estimation with Application to Tracking and Navigation. Wiley-Interscience.
31. Dogandzic A & Nehorai A (2001) Cramer-Rao bounds for estimating range,

velocity, and direction with an active array. IEEE Transactions on Signal Processing 49(6): 1122–1137.
32. Stoica P & Moses RL (1997) Introduction to Spectral Analysis. Englewood Cliffs, New Jersey, Prentice Hall.
33. Chen H (2011) Accurate and efficient node localization in wireless sensor networks. Ph.D. thesis, The University of Tokyo.
34. Gholami MR, Wymeersch H, Ström E & Rydström M (2011) Wireless network positioning as a convex feasibility problem. Eurasip special issue on Localization in Mobile Wireless and Sensor Networks.
35. Venkatesh S & Buehrer RM (2007) NLOS mitigation using linear programming in ultrawideband location-aware networks. IEEE Transaction on Vehicular Technology 56(5): 3182–3198.
36. Gower JC (1966) Some distance properties of latent root and vector methods in multivariate analysis. Biometrika 53: 315–328.
37. Cox TF & Cox MAA (2000) Multidimensional Scaling. 2 ed. Chapman & Hall/CRC.
38. Borg I & Groenen P (1997) Modern Multidimensional Scaling. Series in Statistics. Springer.
39. Cayton L & Dasgupta S (2006) Robust Euclidean embedding. In: on Machine Learning (ICML) International conference on Machine Learning.
40. Dattorro J (2005) Convex Optimization and Euclidean Distance Geometry. Meboo Publishing.
41. Gower JC (1985) Properties of Euclidean and non-Euclidean distance matrices. Linear Algebra and its Applications 67: 81–97.
42. So AMC & Ye Y (2005) Theory of semidefinite programming for sensor network localization. In: ACM Proceedings of the 16^{th} Annual ACM-SIAM Symposium on Discrete Algorithms: 405–414.
43. Chan FKW, So HC & Ma WK (2009) A novel subspace approach for cooperative localization in wireless sensor networks using range measurements. IEEE Transactions on Signal Processing 57(1): 260–269.
44. Latsoudas G & Sidiropoulos ND (2007) A fast and effective multidimensional scaling approach for node localization in wireless sensor networks. IEEE Transactions on Signal Processing 55(10): 5121–5127.
45. Cheung KW & So HC (2005) A multidimensional scaling framework for mobile location using time-of-arrival measurements. IEEE Transactions on Signal Processing 53(2): 460–470.
46. Mardia KV (1978) Some properties of classical multi-dimensioinal scaling. Communicaitons in Statistics - Theory and Methods: 1233–1241.
47. Trosset MW (2002) Extensions of classical multidimensional scaling: Computational theory. Computational Statistics 17.
48. Alfakih AY & Wolkowicz H (1998) On the embeddability of weighted graphs in Euclidean spaces. Technical report, University of Waterloo, Dept. of Combinatorics and Optimization.
49. Graepel T (2002) Kernel matrix completion by semidefinite programming. In: Dorronsoro JR (ed) International Conference on Artificial Neural Networks, volume Vol. 2415: 694–699.

50. Alfakih AY, Wolkowicz H & Khandani A (1999) Solving Euclidean distance matrix completion problems via semidefinite programming. Journal on Computational Optimization and Applications 12(1): 13–30. URI: `http://portal.acm.org/citation.cfm?id=316254#`.
51. Biswas P, Liang TC, Toh KC, Wang TC & Ye Y (2006) Semidefinite programming approaches for sensor network localization with noisy distance measurements. In: IEEE Transactions on Automation Science and Engineering, volume 3 of *1545-5955*: 360–371.
52. Biswas P, Liang TC, Toh KC & Wang TC (2005) Semidefinite programming based algorithms for sensor network localization with noisy distance measurements. Technical report, Dept. of Management Science Engineering, Stanford University.
53. Lui K, Chan F & So H (2009) Semidefinite programming approach for range-difference based source localization. IEEE Transactions on Signal Processing 57(4): 1630–1633.
54. Xu E, Ding Z & Dasgupta S (2011) Source localization in wireless sensor networks from signal time-of-arrival measurements. IEEE Transactions on Signal Processing 59(6): 2887–2897.
55. Boyd S & Vandenberghe L (2004) Convex Optimization. Cambridge University Press.
56. Krislock N & Wolkowicz H (2010) Explicit sensor network localization using semidefinite representations and facial reductions. SIAM J. Optim. 20: 2679–2708.
57. Kim S, Kojima M & Waki H (2009) Exploiting sparsity in SDP relazation for sensor network localization. SIAM J. Optim. 20(1): 192–215.
58. Nocedal J & Wright SJ (2006) Numerical Optimization, volume XXII. Springer, 2 edition.
59. Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM 11(2): 431–441.
60. Kuo SP & Tseng YC (2011) Discriminant minimization search for large-scale RF-based localization systems. IEEE Transactions on Mobile Computing 10(2): 291–304.
61. Destino G & de Abreu GTF (2006) Sensor localization from WLS optmization with closed-form gradient and Hessian. In: Proc. IEEE Global Telecommunications Conference: 1–6.
62. Calafiore G, Carlone L & Mingzhu W (2010) A distributed gradient method for localization of formations using relative range measurements. In: IEEE International Symposium on Computer-Aided Control System Design: 1146–1151.
63. Zhu S & Ding Z (2011) Distributed cooperative localization of wireless sensor networks with convex hull constraint. IEEE Transaction on Wireless Communications 10(7): 2150–2161.
64. Abramo A, Blanchini F, Geretti L & Savorgnan C (2008) A mixed convex/nonconvex distributed localization approach for the deployment of indoor positioning services. IEEE Transactions on Mobile Computing 7(11): 1325–1337.
65. A Beck PS & Li J (2008) Exact and approximate solutions of source localization problems. IEEE Transaction on Signal Processing 56: 1770–1778.
66. Larsson EG & Danev D (2010) Accuracy comparison of LS and squared-range LS for source localization. IEEE Transaction on Signal Processing 58(2): 916–923.

67. de Leeuw J (2000) Multidimensional scaling. Technical report.
68. Groenen PJF (1993) The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions. DSWO Press, Leiden University.
69. Patwari N, Ash J, Kyperountas S, Moses R & Correal N (2005) Locating the nodes: Cooperative localization in wireless sensor networks. IEEE Signal Processing Magazine: 54–69.
70. Patwari N, Dea RJO & Wang Y (2001) Relative location in wireless networks. In: Vehicular Technology Conference, volume 2: 1149–1153.
71. Ali MM, Törn A & Viitanen S (1997) A numerical comparison of some modified controlled random search algorithms. Journal of Global Optimization 11(4).
72. Yao Y (1989) Dynamic tunneling algorithm for global optimization. IEEE Transactions on Systems, Man and Cybernetics 19(5).
73. Destino G & Abreu G (2011) On the maximum likelihood solution of source and network localization. IEEE Transactions on Signal Processing 59(10): 4954–4970.
74. Fidan B, Dasgupta S & Anderson BDO (2008) Guaranteeing practical convergence in algorithms for sensor and source localization. IEEE Transaction on Signal Processing 56(9): 4458–4469.
75. Peterka V (1981) Bayesian approach to system identification. In Eykhoff P (ed) Trends and progress in system identification. Oxford Pergamon Press: 219–304.
76. Ihler A, Fisher J, Moses R & Willsky A (2005) Nonparametric belief propagation for self-localization of sensor networks. IEEE Journal on Selected Areas in Communications 23(4): 809–819.
77. Ristic B, Arulampalam S & Gordon N (2004) Beyond the Kalman Filter: Particles Filters for Tracking Applications. Artech House.
78. Kalman RE (1960) A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering 82(Series D): 35–45.
79. Alspach DL & Sorenson HW (1972) Nonlinear Bayesian estimation using Gaussian sum approximations. IEEE Transactions on Automatic Control 4: 439–448.
80. Williams JL (2003) Gaussian mixture reduction for tracking multiple maneuvering tragets in clutter. Master's thesis, Air Force Inst. of Tech. school of engineering and management.
81. Löytty SA (2009) Gaussian mixture filters in hybrid positioning. Ph.D. thesis, Tampere University of Technology.
82. Bilik I & Tabrikian J (2010) Maneuvering target tracking in the presence of glint using the nonlinear Gaussian mixture Kalman filter. IEEE Transaction on Aerospace and Electronic Systems 46(1): 246–262.
83. Julier SJ & Uhlmann JK (2004) Unscented filtering and nonlinear estimation. IEEE Proceedings 92(3): 401–422.
84. Arasaratnam I & Haykin S (2009) Cubature Kalman filter. IEEE Transactions on Automatic Control 54(6): 1254–1269.
85. Pradhan SS, Kusuma J & Ramchandran K (2002) Distributed compression in a dense microsensor network. IEEE Signal Processing Magazine 19(2): 51–60.
86. Arulampalam MS, Maskell S, Gordon N & Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transaction on Signal Processing 50(2): 174–188.

87. Daum F (2005) Nonlinear filters: beyond the Kalman filter. IEEE Aerospace and Eletronic Systems Magazine 20(8): 57–69.
88. Doucet A, de Freitas N & Gordon NJ (2001) Sequential Monte Carlo methods in Practise. New York: Srpinger.
89. Gordon NJ, Salmond DJ & Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceeding F, Radar and Signal Processing 140(2): 107–113.
90. Olfati-Saber R (2007) Distributed Kalman filtering for sensor networks. In: IEEE Conference on Decision and Control: 5492–5498.
91. Olfati-Saber R (2007) Distributed tracking for mobile sensor networks with information-driven mobility. In: American Control Conference: 4606–4612.
92. Kar S & Moura J (2011) Gossip and distributed Kalman filtering: Weak consensus under weak detectability. IEEE Transaction on Signal Processing 59(4): 1766–1784.
93. Wymeersch H, Ferner U & Win MZ (2008) Cooperative Bayesian self-tracking for wireless networks. IEEE Communications Letter 12(7).
94. Wymeersch H, Lien J & Win M (2009) Cooperative localization in wireless networks. IEEE Proceedings 97(2): 427–450.
95. Olfati-Saber R, Fax J & Murray R (2007) Consensus and cooperation in networked multi-agent systems. IEEE Proceedings 95(1): 215–233.
96. Bickson D (2009) Gaussian belief propagation: Theory and application. Ph.D. thesis, The Hebrew University of Jerusalem.
97. Wymeersch H (2007) Iterative Receiver Design. Cambridge University Press.
98. Carter G (1981) Time delay estimation for passive sonar signal processing. IEEE Transactions on Acoustics, Speech and Signal Processing 29(3): 463–470.
99. Knight W, Pridham R & Kay S (1981) Digital signal processing for sonar. IEEE Proceedings 69(11): 1451–1506.
100. Glisson T, Black C & Sage A (1970) On sonar signal analysis. IEEE Transactions on Aerospace and Electronic Systems AES-6(1): 37–50.
101. Pulford G (2005) Taxonomy of multiple target tracking methods. IEE Proceedings.
102. Cox IJ (1993) A review of statistical data association techniques for motion correspondence. International Journal of Computer Vision: 53–66.
103. Mahler RPS (2007) Statistical Multisource-Multitarget Information Fusion. Artech House.
104. Pulford WG (2002) Relative performance of single-scan algorithms for passive sonar tracking. In: Acoustical Soc.
105. Morefield CL (1977) Application of 0-1 integer programming to multitarget tracking problems. IEEE Transactions on Automatic Control 22(3).
106. Reid D (1979) An algorithm for tracking multiple targets. IEEE Transactions on Automatic Control 24(6): 843–854.
107. Blackman SS (2004) Multiple hypothesis tracking for multiple target tracking. IEEE Transactions on Aerospace and Electronic Systems 19(1): 5–18.
108. Davey SJ (2003) Extensions to the probabilistic multi-hypothesis tracker for improved data association. Ph.D. thesis, The University of Adelaide.
109. Yanhua R & Willett P (2001) An improved PMHT using an idea from coding. In: Aerospace Conference, volume 4: 1889–1900.

110. Bar-Shalom Y (1975) Tracking in a cluttered environment with probabilistic data association. Automatica.
111. Musicki D, Evans R & Stankovic S (1994) Integrated probabilistic data association. IEEE Transaction on Automatic Control 39(6): 1237–1241.
112. Chakravorty R & Challa S (2009) Multitarget tracking algorithm - joint IPDA and Gaussian mixture PHD filter. In: International Conference on Information Fusion: 316–123.
113. Tobias M (2007) Probability hypothesis density for multitarget multisensor tracking with application to passive radar. Ph.D. thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology.
114. Vihola M (2004). Random sets for multitargets tracking and data fusion. LIc. Thesis.
115. Panta K (2007) Multi-target tracking using 1st moment of random finite sets. Ph.D. thesis, Department of Electrical and Electronic Engineering, The University of Melbourne.
116. Mahler R (2003) Multitarget Bayes filtering via first-order multitarget moments. IEEE Transactions on Aerospace and Electronic Systems 39(4): 1152–1177.
117. Vo BN & Doucet A (2005) Sequential Monte Carlo methods for multitarget filtering with random finite sets. IEEE Transactions on Aerospace and Electronic Systems 41(4): 1224–1245.
118. Vo BN & Ma WK (2006) The Gaussian mixture probability hypothesis density filter. IEEE Transactions on Signal Processing 54(11): 4091–4104.
119. Vo BT, Vo BN & Cantoni A (2007) Analytic implementation of the cardinalized probability hypothesis density filter. IEEE Transactions Signal Processing 55(7): 3553–3567.
120. Clark DE (2006) Multiple target tracking with the probability hypothesis density filter. Ph.D. thesis, Heriot-Watt University, Department of Electrical, Electronic and Computing Engineering.
121. Vo BT (2008) Random finite sets in multi-objects filtering. Ph.D. thesis, School of Electrical, Electronic and Computer Engineering.
122. Schoenberg IJ (1935) Remarks to Maurice Frechet's article "Sur la definition axiomatique d'une classe d'espace distances vectoriellement applicable sur l'espace de hilbert. The Annals of Mathematics 36(3): 724–732.
123. Sibson R (1979) Studies on the robustness of multidimensional scaling: Perturbation analysis of classical scaling. Journal of the Royal Statistical Society. Series B 41(2): 217–229.
124. Meulman JJ (1992) The integration of multidimensional scaling and multivariate analysis with optimal transformations. Psychometrika 57.
125. Kearsley AJ, Tapia RA & Trosset MW (1998) The solution of the metric stress and sstress problems in multidimensional scaling using newton's method. Computational Statistics 13(3): 369–396.
126. Kruskal J (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29: 1–27.
127. Takane Y, Young F & de Leeuw J (1984) Nonmetric individual differences in multidimensional scaling: An alternating least squares method with optimal transformation. Psychometrika 42: 7–67.

128. Cheung KW, So HC, Ma WK & Chan YT (2004) Least squares algorithms for time-of-arrival-based mobile location. IEEE Transaction on Signal Processing 52(4): 1121–1128.
129. Chen ZX, Wei HW, Wan Q, Ye SF & Yang WL (2009) A supplement to multidimensional scaling framework for mobile location. IEEE Transactions on Signal Processing 57(5).
130. de Leeuw J (1977) Applications of Convex Analysis to Multidimensional Scaling. Recent Developments in Statistics. North Holland Publishing Company.
131. MM Bronstein RK AM Bronstein & Yavneh I (2006) Multigrid multidimensional scaling. Numer. Linear Algebra Appl. 13: 149–171.
132. Rosman G, Bronstein AM, Bronstein MM & Kimmel R (2008). Fast multidimensional scaling using vector extrapolation. Techn. Report CIS-2008-01, Dept. of Computer Science.
133. Jazwinski AH (1970) Stochastic Processes and Filtering Theory. Dover Books on Electrical Engineering.
134. Särkkä S (2006) Recursive Bayesian inference on stochastic differential equations. Ph.D. thesis, Helsinki University of Technology.
135. Mahler R (2007) PHD filters of higher order in target number. IEEE Transactions on Aerospace and Electronic Systems 43(3): 1523–1543.
136. Vo BT, Vo BN & Cantoni A (2007) Analytic implementation of the cardinalized probability hypothesis density filter. IEEE Transactions on Signal Processing 55(7): 3553–3567.
137. Erdnic O, Willett P & Bar-Shalom Y (2009) The bin-occupancy filter and its connection to the PHD filter. IEEE Transactions Signal Processing 57(11): 4232–4246.
138. Hernandez S (2010) Smoothing algorithms for the probability hypothesis density filter. Technical report, Victoria University of Wellington.
139. Schuhmacher D, Vo BT & Vo BN (2008) A consistent metric for performance evaluation of multi-object filters. IEEE Transactions on Signal Processing 56(8): 3447–3457.
140. Yang B (1995) Projection approximation subspace tracking. IEEE Transactions on Signal Processing 43(1): 95–107.
141. Xu G & Kailath T (1994) Fast subspace decomposition. IEEE Transactions on Signal Processing 42(3): 539–551.
142. Comon P & Golub GH (1990) Tracking a few extreme singular values and vectors in signal processing. IEEE Proceedings: 1327–1343.
143. Golub GH & van Loan CF (1996) Matrix Computations. Johns Hopkins Univ. Press, New York, NY, 3 edition.
144. Torgerson W (1952) Multidimensional scaling 1: Theory and method. Psychometrika 17: 401–419.
145. Tulino AM & Verdu S (2004) Random Matrix Theory And Wireless Communications. Foundations and Trends in Communications and Information Theory. Now Publishers Inc.
146. Williams C & Seeger M (2000) Using the Nyström method to speed up kernel machines. Annual Advances in Neural Information Processing Systems 13: 682–688.

147. Schölkopf B, Smola A & Muller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10(1299-1319).
148. Fowlkes C, Belongie S, Chung F & Malik J (2004) Spectral grouping using the Nyström method. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(2): 214–225.
149. Gershgorin SA (1931) Über die Abgrenzung der Eigenwerte einer matrix. Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. 7: 749–754.
150. Nobakhti M & Munro N (2004) A new method for singular value loop shaping in design of multiple-channel controllers. IEEE Transactions on Automatic Control 49(2): 249–253.
151. Varga RS (2004) Gershgorin and His Circles. Computational Mathematics. Springer.
152. Andrews GE, Askey R & Roy R (1999) Special Functions: Encyclopedia of Mathematics. Cambridge University Press, Cambridge, UK, 1 edition.
153. Filho JCSS & Yacoub M (2004) Nakagami-m approximation to the sum of m non-identical independent Nakagami-m variates. IEE Electronic Letters 40(15).
154. David HA & Nagaraja HN (2003) Order Statistics. Series in Probability and Statistics. Wiley, third edition.
155. Song G & Li YG (2006) Asymptotic throughput analysis for channel-aware scheduling. IEEE Transaction on Wireless Communications 54(10): 1827–1834.
156. Parlett BN (1980) The Symmetric Eigenvalue Problem. Series in Computational Mathematics. Prentice Hall.
157. Cardoso JF & Souloumiac A (1996) Jacobi angles for simultaneous diagonalization. SIAM Journal on Matrix Analysis and Applications 17(1): 161–164.
158. Bunse-Gerstner A, Byers R & Mehrmann V (1993) Numerical methods for simultaneous diagonalization. SIAM Journal on Matrix Analysis and Applications 14(4): 927–949.
159. Hua Y, Xianga Y, Chen T, Abed-Meraim K & Miao Y (1999) A new look at the power method for fast subspace tracking. Digital Signal Processing 9(4): 297–314.
160. Badeau R, David B & Richard G (2005) Fast approximated power iteration subspace tracking. IEEE Transaction on Signal Processing 53(8): 2931–2941.
161. Demmel J & Veselic K (1992) Jacobi's method is more accurate than QR. SIAM J. Matrix Anal. Appl.
162. Drmac Z & Veselic K (2005) New fast and accurate Jacobi SVD algorithm: I-II. Technical report, Linear Algebra PACKage (LAPACK), http://www.netlib.org/lapack/lawns/downloads/.
163. Holobar A, Ojstersek M & Zazula D (2006) Distributed Jacobi joint diagonalization on clusters of personal computers. International Journal of Parallel Programming 34(6).
164. Drmac Z (2007) A global convergence proof of cyclic Jacobi methods with block rotations. Technical report, Linear Algebra PACKage (LAPACK), http://www.netlib.org/lapack/lawns/downloads/.
165. Shen H & Huper K (2009) Block Jacobi-type methods for non-orthogonal joint diagonalization. In: IEEE International Conference on Acoustics, Speech and Signal Processing: 3285–3288.

166. Higham NJ (2002) Accuracy and Stability of Numerical Algorithms. 2 ed. Philadelphia PA, SIAM.
167. Macagnano D & de Abreu GTF (2008) Tracking multiple dynamic targets in LOS-NLOS condition with multidimensional scaling. In: IEEE Workshop on Positioning, Navigation and Communication.
168. SMallat (1998) A Wavelet Tour of Signal Processing. 2 ed. Academic Press.
169. Platt JC (2004) Fastmap, metricmap, and landmark mds are all nyström algorithms. Technical Report MSR-TR-2004-26, Microsoft Research Technical Report. URI: `http://citeseer.ist.psu.edu/732858.html`.
170. Shang Y & Ruml W (2004) Improved MDS-based localization. In: Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom'04), volume 4: 2640–2651.
171. Yu K & Guo Y (2008) Improved positioning algorithms for nonline-of-sight environments. IEEE Transaction on Vehicular Technology 57(4): 2342–2353.
172. Mourad F, Snoussi H, Abdallah F & Richard C (2009) Anchor-based localization via interval analysis for mobile ad-hoc sensor networks. IEEE Transactions on Signal Processing 57(8): 3226–3239.
173. de Abreu GTF (2008) On the generation of Tikhonov variates. IEEE Transactions on Communications 56(7): 1157–1168.
174. Julier SJ & Uhlmann JK (1997) A consistent debiased method for converting between polar and Cartesian coordinates systems. In: AeroSense: Acquisition, Tracking and Pointing: 110–121.
175. Julier SJ (1998) A skewed approach to filtering. In: AeroSense: Aerospace/Defence, Simulation and Controls: 182–193.
176. der Merwe RV & Wan EA (2001) The square-root unscented Kalman filter for state and parameter-estimation. In: IEEE International Conference on Acoustics, Speech and Signal Processing, volume 6: 3461–3464.
177. Wu Y, Hu D, Wu M & Hu X (2006) A numerical-integration perspective on Gaussian filters. IEEE Transactions on Signal Processing 54(8): 2910–2921.
178. Daley D & Vere-Jones D (2009) An Introduction to the Theory of Point Processes. Springer-Verlag.
179. Vieta. URI: `http://mathworld.wolfram.com/VietasFormulas.html`.
180. Ulmke M, Erdnic O & Willett P (2010) GMTI tracking via the Gaussian mixture cardinalized probability hypothesis density filter. IEEE Aerospace and Eletronic Systems Magazine 46(4): 1821–1833.
181. Bailey T, Upcroft B & Durrant-Whyte H (2006) Validation gating for non-linear non-Gaussian target tracking. In: IEEE Conference on Information Fusion.
182. Breidt FJ & Carriquiry AL (2000) Highest density gates for target tracking. IEEE Aerospace and Eletronic Systems Magazine 36(1): 47–55.
183. Vo BT, Vo BN & Cantoni A (2009) The cardinalized balanced multi-target multi-bernoulli filter and its implementation. IEEE Transaction on Signal Processing 57(2): 409–423.
184. Murga DS & Iserte AP (2010) Differential feedback of MIMO channel Gram matrices based on geodesic curves. IEEE Transaction on Wireless Communications 9(12): 3714–3727.

185. Mahler RPS, Vo BT & Vo BN (2011) CPHD filtering with unknown clutter rate and detection profile. IEEE Transaction on Signal Processing (pre-print) 59(8): 3497–3513.
186. Prudnikov A, Brychkov Y & Marichev O (2003) Integrals and Series - Elementary Functions. (in Russian). Moscow Fizmatlit.
187. Proakis JG (2000) Digital Communications, Fourth Edition. McGraw-Hill, New York, NY.
188. Abramowitz M & Stegun IA (1972) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, 10 edition.
189. de Abreu GTF (2010) On the moment-determinance and random mixture of Nakagami-m variates. IEEE Transactions on Communications 58(9): 2561–2575.

Appendix 1: Procrustes analysis

As mentioned in Section 3.1 the methods solving the scaling problem, *e.g.* the C-MDS, find a configuration of points that is *geometrically congruent* to the true one up to an *affine transformation*, namely a rigid transformation[88] and a scaling factor. To remove this uncertainty in the solution the Procrustes transformation detailed in the following must be utilized [37].

Assuming that the coordinates of at least $N_A > \eta$ anchor nodes $\mathbf{X}_{N_A} \in \mathbb{R}^{N_A \times \eta}$ are known, then the final estimate $\hat{\mathbf{X}}$ can be recovered from the output $\mathbf{Y} \in \mathbb{R}^{N \times \eta}$ of the scaling algorithm using the following isotropic dilation and the rigid transformations

$$\hat{\mathbf{X}} = \alpha \cdot \mathbf{Y} \cdot \mathbf{F}_{\mathrm{R}} + \mathbf{c}_{\mathrm{D}} \otimes \mathbf{1}_{[N \times 1]}, \tag{238}$$

where the scalar α, the matrix $\mathbf{F}_{\mathrm{R}} \in \mathbb{R}^{\eta \times \eta}$ and the vector $\mathbf{c}_{\mathrm{D}} \in \mathbb{R}^{\eta}$ are the coefficients of the Procrustes transformation, calculated as

$$\alpha = \mathrm{tr}\Big([\mathbf{H}_{N_A}^{\mathrm{T}} \cdot \mathbf{H}_{N_A}]^{\frac{1}{2}}\Big) \cdot \frac{\|\bar{\mathbf{X}}_{N_A}\|_{\mathrm{F}}}{\|\bar{\mathbf{Y}}_{N_A}\|_{\mathrm{F}}}, \tag{239}$$

$$\mathbf{F}_{\mathrm{R}} = \mathbf{U}_{\mathrm{L}}^{\mathrm{T}} \cdot \mathbf{U}_{\mathrm{R}}, \tag{240}$$

$$\mathbf{c}_{\mathrm{D}} = \frac{\mathbf{1}_{N_A}^{\mathrm{T}}}{N_A} \cdot \mathbf{X}_a - \frac{\alpha \cdot \mathbf{1}_{N_A}^{\mathrm{T}}}{N_A} \cdot [\mathbf{Y}]_{1:N_A,1:\eta} \cdot \mathbf{F}_{\mathrm{R}}. \tag{241}$$

In the above, $\mathbf{U}_{\mathrm{L}}$ - $\mathbf{U}_{\mathrm{R}}$ are the left and right matrices of singular vectors of $\mathbf{H}_{N_A} = \bar{\mathbf{X}}_{N_A}^{\mathrm{T}} \cdot \bar{\mathbf{Y}}_{N_A} = \mathbf{U}_{\mathrm{L}} \cdot \mathbf{\Sigma} \cdot \mathbf{U}_{\mathrm{R}}$, and $\bar{\mathbf{X}}_{N_A}$ and $\bar{\mathbf{Y}}_{N_A}$ are the centralized true and rotated coordinate matrices of the anchor nodes, given by

$$\bar{\mathbf{X}}_{N_A} = \mathbf{X}_{N_A} - \frac{\mathbf{1}_{N_A}^{\mathrm{T}}}{N_A} \cdot \mathbf{X}_{N_A} \otimes \mathbf{1}_{N_A}, \tag{242}$$

$$\bar{\mathbf{Y}}_{N_A} = \mathbf{Y}_{N_A} - \frac{\mathbf{1}_{N_A}^{\mathrm{T}}}{N_A} \cdot [\mathbf{Y}]_{1:N_A,1:\eta} \otimes \mathbf{1}_{N_A}. \tag{243}$$

[88] Rigid transformations are reflections, rotations and translations [37, 38].

Appendix 2: Proof of Lemma 3

First, recognize that $\varepsilon_i = a_{ii} \sim p_{\text{Gamma}}(r; \eta/2, 2\sigma_y^2)$, which also implies that ε_i are i.i.d. with respect to the subindex i which therefore will be temporarily dropped from the notation, for simplicity.

Given the distributions of ϵ and ε, it is clear that

$$\Pr\{\epsilon \geqslant \varepsilon \mid \varepsilon\} = \frac{2}{R_\epsilon} \int_{\varepsilon}^{\infty} r \cdot \exp\left(\frac{-r^2}{R_\epsilon}\right) \mathrm{d}r = \exp\left(\frac{-\varepsilon^2}{R_\epsilon}\right). \tag{244}$$

Marginalizing equation (244) over the distribution of ε yields

$$\begin{aligned} \Pr\{\epsilon \geqslant \varepsilon\} &= \int_0^{\infty} \exp\left(\frac{-r^2}{R_\epsilon}\right) \cdot p_{\text{Gamma}}\left(r; \frac{\eta}{2}, 2\sigma_y^2\right) \mathrm{d}r \qquad (245) \\ &= \frac{1}{(2\sigma_y^2)^{\eta/2}\,\Gamma(\frac{\eta}{2})} \int_0^{\infty} r^{\eta/2-1} \cdot \exp\left(-\frac{r^2}{R_\epsilon} - \frac{r}{2\sigma_y^2}\right) \mathrm{d}r. \end{aligned}$$

Using the closed-form solution to the latter integral[89] in [186, pp.297, Eq.(9)], we obtain

$$\Pr\{\epsilon \geqslant \varepsilon\} = \left(\frac{R_\epsilon}{8\sigma_y^4}\right)^{\frac{\eta}{4}} \cdot \exp\left(\frac{R_\epsilon}{32\sigma_y^4}\right) \cdot D_{-\eta/2}\left(\sqrt{\frac{R_\epsilon}{8\sigma_y^4}}\right). \tag{246}$$

Finally, equation (246) reduces to (96) by use of the relation [186, pp.621]

$$D_\nu(r) = 2^{\frac{\nu}{2}} \exp(-r^2/4) U(-\nu/2, 1/2, r^2/2), \tag{247}$$

which concludes the proof.

[89] Setting $c = 0$ in the cosh variation of [186, pp.297, Eq.(9)] yields $\int_0^\infty \exp(-ax^2 - bx) x^{\alpha-1} \mathrm{d}x = \Gamma(\alpha) \cdot (2a)^{-\alpha/2} \exp(b^2/(8a)) \cdot D_{-\alpha}(b/\sqrt{2a})$, where $D_\nu(x)$ is the Parabolic Cylinder function.

Appendix 3: Proof Lemma 4

First, notice that the η-tuple $\mathbf{h}_i \triangleq \{h_{i1}, \cdots, h_{i\eta}\}$ is constant for any given i. Let us define the variable $\zeta_{ij}|\mathbf{h}_i \triangleq |\sum_{k=1}^{\eta} h_{ik} \cdot w_{jk}|$. Obviously $\zeta_{ij}|\mathbf{h}_i$'s are half-Normal or, alternatively,

$$\zeta_{ij}|\mathbf{h}_i \sim p_{\text{Nakagami}}(r, 0.5, \sigma^2 \sum_{k=1}^{\eta} h_{ik}^2), \tag{248}$$

where

$$p_{\text{Nakagami}}(r; m, R) \triangleq \frac{2r^{2m-1}}{\Gamma(m)} \cdot \left(\frac{m}{R}\right)^m \cdot \exp\left(-\frac{m}{R} r^2\right). \tag{249}$$

For notational simplicity we shall hereafter omit the conditional dependence of $\zeta_{ij}|\mathbf{h}_i$ on $\mathbf{h}_i$ and write simply ζ_{ij}.

Consequently, the n-th moment of ζ_{ij}, with respect to realizations over j, is [187, pp. 47],

$$\mathrm{E}[\zeta_{ij}^n] = \frac{1}{\sqrt{\pi}} \cdot \Gamma\left(\frac{n+1}{2}\right) \cdot \left(\sqrt{2\sigma^2 \cdot \sum_{k=1}^{\eta} h_{ik}^2}\right)^n = M_n(1/2, 1/2) \cdot F_i^{\,n}, \tag{250}$$

where the function[90] $\mathcal{M}_n(m, R)$ and the auxiliary quantity F_i are defined below:

$$\mathcal{M}_n(m, R) \triangleq \frac{\Gamma(m + n/2)}{\Gamma(m)} \cdot \left(\frac{R}{m}\right)^{\frac{n}{2}}, \tag{251}$$

$$F_i \triangleq \sqrt{2\sigma^2 \cdot \sum_{k=1}^{\eta} h_{ik}^2}. \tag{252}$$

In light of the above, and by force of the multinomial theorem and the independence[91] of ζ_{ij} , the n-th moments of z_i with respect to different realizations of h_{ik} and y_{kj}, is

$$\begin{aligned} \mathrm{E}[z_i^n] &= \mathrm{E}\left[\left(\sum_{j=1}^{M} \zeta_{ij}\right)^n\right] = \sum_{\{q_1, \cdots, q_M\}} n! \cdot \frac{\mathrm{E}[\zeta_{ij}^{q_1} \cdots \zeta_{ij}^{q_M}]}{q_1! \cdots q_M!} \\ &= n! \cdot \sum_{\{q_1, \cdots, q_M\}} \mathrm{E}\left[\prod_{j=1}^{M} \frac{\mathcal{M}_{q_j}(1/2, 1/2) \cdot F_i^{q_j}}{q_j!}\right], \end{aligned} \tag{253}$$

[90] Notice that $\mathcal{M}_n(m, R)$ is simply the envelope moment function of the Nakagami-m pdf.

[91] To clarify, ζ_{ij} are independent amongst themselves but conditionally dependent on $\mathbf{h}_i$. Consequently, in equation (253), the Law of Iterated Expectations is implicitly invoked.

where the summations in the two rightmost equalities are taken over all M-tuples $\{q_1, \cdots, q_M\}$ of non-negative integers satisfying the Diophantine equation $q_1 + \cdots + q_M = n$, hereafter referred to as the *Diophantine constraint*.

Next, realize that due to the Diophantine constraint $\prod_{j=1}^{M} F_i^{q_j} = F_i^{\sum_{j=1}^{M} q_j} = F_i^n$. Furthermore, $\mathcal{M}_n(1/2, 1/2) = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{n+1}{2}\right)$, such that equation (253) reduces to

$$\mathrm{E}[z_i^n] = n! \cdot \mathrm{E}[F_i^n] \cdot \sum_{\{q_1, \cdots, q_M\}} \prod_{q_j \neq 0} \frac{1}{q_j} \cdot \frac{\Gamma\left(\frac{q_j}{2} + \frac{1}{2}\right)}{\sqrt{\pi} \cdot \Gamma(q_j)}, \tag{254}$$

where we have used the relation $r! = \Gamma(r+1) = r \cdot \Gamma(r)$ and the last product is taken over the non-zero integers q_k of the corresponding Diophantine-constrained M-tuple $\{q_1, \cdots, q_M\}$.

Using the Legendre duplication formula for the Gamma function [188, pp. 256, eq. 6.1.18] in the product appearing in equation (254) yields

$$\mathrm{E}[z_i^n] = n! \cdot \mathrm{E}[F_i^n] \cdot \sum_{\{q_1, \cdots, q_M\}} \prod_{q_j \neq 0} \frac{1}{2^{q_j} \cdot \Gamma\left(\frac{q_j}{2} + 1\right)}. \tag{255}$$

Due to the monotonically ascending behavior of the function $2^r\Gamma(r/2+1)$ on $r \geqslant 1$, the denominator in the product appearing in equation (255) is smallest for $q_j = 1, \forall\, j$. Consequently $\min\limits_{q_j}\{2^{q_j} \cdot \Gamma\left(\frac{q_j}{2} + 1\right)\} = 2^1 \cdot \Gamma\left(\frac{1}{2} + 1\right) = \sqrt{\pi}$.

In other words, the dominant terms of the sum over all distinct M-tuple $\{q_1, \cdots, q_M\}$ satisfying the Diophantine constraint are the ones associated with the M-tuple in which all q_j's are either 0 or 1, exclusively. Furthermore, in each such M-tuple, there are exactly n distinct q_j's equal to one, such that the largest terms in the sum are $\pi^{-n/2}$.

Next, realize that there are $\binom{M}{n}$ as many such (dominant) terms, which yields

$$\sum_{\{q_1, \cdots, q_M\}} \prod_{q_j \neq 0} \frac{1}{2^{q_j} \cdot \Gamma\left(\frac{q_j}{2}+1\right)} \underset{M \gg 1}{\longrightarrow} \binom{M}{n} \frac{1}{\pi^{n/2}}. \tag{256}$$

Furthermore, expanding the binomial coefficient and considering its largest term we obtain

$$\binom{M}{n} = M(M-1)(M-2)(M-n-1) \underset{M \gg 1}{\longrightarrow} M^n. \tag{257}$$

Substituting equations (252), (256) and (257) into equation (255), we finally

obtain for a sufficiently large M

$$\mathrm{E}[z_i^n] \underset{M \gg 1}{\longrightarrow} \left(\frac{2M^2\sigma^4}{\pi}\right)^{\frac{n}{2}} \cdot \mathrm{E}\left[\left(\sqrt{\frac{1}{\sigma^2}\sum_{k=1}^{\eta} h_{ik}^2}\right)^n\right] = \left(\sqrt{\alpha/2}\right)^n \cdot \mathrm{E}[\chi_i^n],\ \forall\, n \in \mathbb{N}, \quad (258)$$

where the σ^2 dividing the sum inside the expectation is maintained explicitly for future convenience, α is as defined in equation (107) and χ_i is a random variate with distributions

$$\chi_i \sim \begin{cases} p_{\mathrm{C}\chi}(r;\eta) \triangleq \dfrac{r^{\eta-1}}{2^{\eta/2-1}\cdot\Gamma(\eta/2)} \cdot \exp\left(\dfrac{-r^2}{2}\right) & \text{if } \mathrm{E}[h_{ik}] = 0\,\forall\, k, \\ p_{\mathrm{NC}\chi}(r;\eta,\beta_i) \triangleq \dfrac{r^{\eta/2}}{\beta_i^{\eta/2-1}} \cdot \exp\left(-\dfrac{r^2+\beta_i^2}{2}\right) \cdot I_{\eta/2-1}(\beta_i r) & \\ & \text{otherwise.} \end{cases} \quad (259)$$

and where $p_{\mathrm{C}\chi}(r;\eta,\alpha)$ and $p_{\mathrm{NC}\chi}(r;\eta,\beta_i,\alpha)$ denote the central and the non-centrale χ distributions.

From equations (258) and (259) it is evident that asymptotically (for large M), all the moments of z_i are identical to the moments of a Chi random variate – central or non-central depending on the means of h_{ik} – multiplied by the constant $\sqrt{\alpha/2}$.

In other words, z_i is[92], a *scaled* Chi variate, which immediately leads[93] to equation (106), concluding the proof.

[92] Rigorously speaking, this assertion implies the moment-determinance of the Chi distribution. This has been proved to be true for the central case in [189], and a similar proof for the non-central case can be easily obtained.

[93] It is well known that if a continuous random variate $w = c \cdot r$, where c is a constant and $r \sim p(r)$, then $w \sim \frac{1}{c} \cdot p(\frac{r}{c})$.

Appendix 4: Proof of Lemma 5

First notice that

$$\phi_{ij} = a_{ij} + b_{ij} + b_{ji} = \sum_{k=1}^{\eta} (w_{ik} \cdot w_{jk} + c_{ik} \cdot w_{jk} + c_{jk} \cdot w_{ik}), \tag{260}$$

$$\theta_{ji} = a_{ij} + b_{ji} = \sum_{k=1}^{\eta} (w_{ik} \cdot w_{jk} + c_{jk} \cdot w_{ik}), \tag{261}$$

where for each realization and the i-th row,

$$\sum_{k=1}^{\eta} w_{ik} \cdot w_{jk} \sim p_{\text{Gauss}}(0; \sigma_y \sqrt{\textstyle\sum_{k=1}^{\eta} w_{ik}^2}), \tag{262}$$

$$\sum_{k=1}^{\eta} c_{ik} \cdot w_{jk} \sim p_{\text{Gauss}}(0; \sigma_y \sqrt{\lambda_i}), \tag{263}$$

$$\sum_{k=1}^{\eta} c_{jk} \cdot w_{ik} \sim p_{\text{Gauss}}(0; \sigma_y \sqrt{\lambda_j}). \tag{264}$$

The relations in equations (262) and (263) are evident, while the one in equation (264) follows from the fact that the η-tuples $\{c_{j1}, \cdots, c_{j\eta}\}$ are orthogonal for different j's.

Since $\hat{\sigma}_y \triangleq \sqrt{\sum_{k=1}^{\eta} w_{ik}^2/\eta}$ is the sample standard deviation of w_{ik}, it is evident that for η sufficiently large $\sqrt{\sum_{k=1}^{\eta} w_{ik}^2} \to \sigma_y\sqrt{\eta}$. It follows that $\phi_{ij} \sim p_{\text{Gauss}}(r; 0, \sigma_y\sqrt{\eta\sigma_y^2 + \lambda_i + \lambda_j})$ and $\theta_{ji} \sim p_{\text{Gauss}}(r; 0, \sigma_y\sqrt{\eta\sigma_y^2 + \lambda_j})$, such that their absolute values are Half-Normal, or alternatively a Nakagami-m variate with $m = 0.5$, which yields equations (111) and (112).

409. Sinisammal, Janne (2011) Työhyvinvoinnin ja työympäristön kokonaisvaltainen kehittäminen – tuloksia osallistuvista tutkimus- ja kehittämisprojekteista sekä asiantuntijahaastatteluista

410. Berg, Markus (2011) Methods for antenna frequency control and user effect compensation in mobile terminals

411. Arvola, Jouko (2011) Reducing industrial use of fossil raw materials : Techno-economic assessment of relevant cases in Northern Finland

412. Okkonen, Jarkko (2011) Groundwater and its response to climate variability and change in cold snow dominated regions in Finland: methods and estimations

413. Anttonen, Antti (2011) Estimation of energy detection thresholds and error probability for amplitude-modulated short-range communication radios

414. Neitola, Marko (2012) Characterizing and minimizing spurious responses in Delta-Sigma modulators

415. Huttunen, Paavo (2012) Spontaneous movements of hands in gradients of weak VHF electromagnetic fields

416. Isoherranen, Ville (2012) Strategy analysis frameworks for strategy orientation and focus

417. Ruuska, Jari (2012) Special measurements and control models for a basic oxygen furnace (BOF)

418. Kropsu-Vehkaperä, Hanna (2012) Enhancing understanding of company-wide product data management in ICT companies

419. Hietakangas, Simo (2012) Design methods and considerations of supply modulated switched RF power amplifiers

420. Davidyuk, Oleg (2012) Automated and interactive composition of ubiquitous applications

421. Suutala, Jaakko (2012) Learning discriminative models from structured multi-sensor data for human context recognition

422. Lorenzo Veiga, Beatriz (422) New network paradigms for future multihop cellular systems

423. Ketonen, Johanna (2012) Equalization and channel estimation algorithms and implementations for cellular MIMO-OFDM downlink

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O.B. 7500 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-951-42-9858-5 (Paperback)
ISBN 978-951-42-9859-2 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO