ANITI-PRAIRIE workshop

26 - 27 June 2023

ENSEEIHT
Toulouse

Welcome to the ANITI-PRAIRIE workshop on Optimization and Machine Learning

This workshop is a joint initiative supported by 3IA institutes from Paris (PRAIRIE) and Toulouse (ANITI). The event is hosted at ENSEEIHT downtown Toulouse. Practical information is found at the end of this booklet.

Scientific commitee:

- Francis Bach (INRIA, Paris)
- Jean-Bernard Lasserre (CNRS, Toulouse)
- Gabriel Peyré (CNRS, Paris)

Local organizing commitee:

- Cédric Févotte (CNRS, Toulouse)
- Edouard Pauwels (Université Paul Sabatier, Toulouse)
- Emmanuel Soubies (CNRS, Toulouse)

Program overview

Workshop ANITI-PRAIRIE		
	Monday 26/6	Tuesday 27/6
09:00 – 10:30	Welcome	Aude Rondepierre Clément Royer
10:30 - 11:00	Introduction	Break
11:00 – 12:30	Didier Henrion Aymeric Dieuleveut	Claire Boyer Jérôme Renault
12:30 - 14:30	Lunch + Poster	
14:30 – 16:00	Valentin De Bortoli Reda Chhaibi	Rachel Bawden Francois Malgouyres
16:00 – 16:30	Break	Goodbye
16:30 – 18:00	Elsa Cazelles Giulio Biroli	
19:30 - …	Gala dinner (Moaï)	

Program: Monday 26

10:30 – 11:00	*Introduction*

11:00 – 11:45 **Didier Henrion** (CNRS, Toulouse):
Polynomial Argmin for Recovery and Approximation of Multivariate Discontinuous Functions

11:45 – 12:30 **Aymeric Dieuleveut** (Ecole Polytechnique, Paris):
Counter-examples in first-order optimization: a constructive approach.

12:30 – 14:30	*Lunch + Posters*

14:30 – 15:15 **Valentin de Bortoli** (CNRS, Paris):
Diffusion Schrödinger Bridge Matching

15:15 – 16:00 **Reda Chhaibi** (Université Paul Sabatier, Toulouse):
Free Probability for predicting the performance fully connected feed-forward neural networks.

16:00 – 16:30	*Break*

16:30 – 17:15 **Elsa Cazelles** (CNRS, Toulouse):
Machine learning and optimal transport: some statistical and algorithmic tools

17:15 – 18:00 **Giulio Biroli** (Ecole Normale Supérieure, Paris):
Renormalisation group and machine learning: the Wavelet-Conditional RG

19:30	*Gala dinner (Moaï, 35 allées Jules Guesde, registration mandatory)*

Program: Tuesday 27

9:00 – 9:45 **Aude Rondepierre** (INSA, Toulouse):
FISTA is an automatic geometrically optimized algorithm for strongly convex functions

9:45 – 10:30 **Clément Royer** (Université Paris Dauphine-PSL):
A Newton-type method for strict saddle functions

10:30 – 11:00	*Break*

11:00 – 11:45 **Claire Boyer** (Sorbonne Université, Paris):
Some statistical insights into PINNs

11:45 – 12:30 **Jérôme Renault** (Toulouse School of Economics):
Optimistic Gradient Descent Ascent in General-Sum Bilinear Games.

12:30 – 14:30	*Lunch + Posters*

14:30 – 15:15 **Rachel Bawden** (INRIA, Paris):
From Linguistic to Visual Context in Machine Translation

15:15 – 16:00 **François Malgouyres** (Université Paul Sabatier, Toulouse):
Deux contributions aux aspects théoriques des réseaux de neurones

16:00 – 16:30	*Goodbye*

Posters

- **Gaspard Beugnot** (INRIA, Paris):
Polynomial optimization with kernel sum-of-squares

- **Mathieu Dagréou** (INRIA, Paris):
A Lower Bound and a Near-Optimal Algorithm for Bilevel Empirical Risk Minimization

- **Valentin DURANTE** (INRAE, Toulouse):
Convex Optimization for Discrete Graphical Models

- **Changqing Fu** (Université Paris Dauphine):
Conic Linear Units: Seamless Model Fusion via Optimal Transport

- **Tam Le** (Toulouse School of Economics):
Nonsmooth implicit differentiation for machine learning and optimization

- **Etienne Lempereur** (ENS, Paris):
Conditionally Strongly Log-Concave Generative Models

- **Jakob Maier** (INRIA, Paris):
Asymmetric tree correlation testing for sparse graph alignment

- **Sibylle Marcotte** (ENS, Paris):
Abide by the law and follow the flow: conservation laws for gradient flows

- **Rachid El Montassir** (CERFACS, Toulouse):
Hybrid Physics-AI architecture for cloud cover nowcasting

- **Zaccharie Ramzi** (ENS, Paris):
Test like you Train in Implicit Deep Learning

- **Antonio Rodrí guez** (Banco Sabadell, Barcelona):
Machine Learning and Regulatory constraints in credit risk

- **Thomas Schiex** (INRAE, Toulouse):
Scalable Coupling of Deep Learning with Logical Reasoning

From Linguistic to Visual Context in Machine Translation

Rachel Bawden

INRIA, Paris

From the very beginning of natural language processing, one of the most important issues has been ambiguity, when words, phrases or sentences have multiple meanings. For Machine Translation (MT), ambiguity is an issue when those meanings result in different translations. Is French trombone a paperclip or a musical instrument? If the fans are out of order, is there a problem at the football (supporteurs) or a ventilation problem (ventilateurs). Traditionally, MT was carried out sentence by sentence, and such sentences could be lacking the context necessary to distinguish between correct and incorrect translations. Since then, contextual MT has been growing in popularity, with the inclusion of both linguistic and extra-linguistic context (e.g. preceding sentences and speaker gender). New evaluation strategies have also been developed to target context-dependent phenomena since the usual metrics are ill adapted. In this talk, I will introduce contextual MT using linguistic context, including some of my earlier work carried out during my PhD, focused on the evaluation of contextual models. I will then describe some of our recent work (Futeral et al., 2023) on the inclusion of visual context for disambiguation in MT, including (i) a new contrastive test set CoMMuTE for the evaluation of the use of visual context and (ii) an adapted multimodal MT approach, VGAMT, which combines pretraining, a visually guided attention mechanism and joint training on two objectives (multimodal MT and visually conditioned masked language modelling). We should that contrarily to previously published methods, VGAMT is able to successfully exploit visual context whilst maintaining MT performance on standard test sets.

Renormalisation group and machine learning: the Wavelet-Conditional RG

Giulio Biroli

Ecole Normale Supérieure, Paris

Reconstructing, or generating, high dimensional distributions starting from data is a central problem in machine learning and data sciences. I will present a method *The Wavelet Conditional Renormalization Group* that combines ideas from physics (renormalization group theory) and computer science (wavelets, stable representations of operators). The Wavelet Conditional Renormalization Group allows to reconstruct in a very efficient way classes of high dimensional probability distributions hierarchically from large to small spatial scales, and to perform RG directly from data. It allows to bridge the gap between approaches based on physical intuition and modern machine learning algorithms. I will present the method and then show its applications to data from statistical physics and cosmology. I shall also discuss the interesting insights that our method offers on the interplay between structures of data and architectures of deep neural networks.

Diffusion Schrödinger Bridge Matching

Valentin de Bortoli

CNRS, Paris

Solving transport problems, i.e. finding a map transporting one given distribution to another, has numerous applications in machine learning. Novel mass transport methods motivated by generative modeling have recently been proposed, e.g. Denoising Diffusion Models (DDMs) and Flow Matching Models (FMMs) implement such a transport through a Stochastic Differential Equation (SDE) or an Ordinary Differential Equation (ODE). However, while it is desirable in many applications to approximate the deterministic dynamic Optimal Transport (OT) map which admits attractive properties, DDMs and FMMs are not guaranteed to provide transports close to the OT map. In contrast, Schrödinger bridges (SBs) compute stochastic dynamic mappings which recover entropy-regularized versions of OT. Unfortunately, existing numerical methods approximating SBs either scale poorly with dimension or accumulate errors across iterations. In this work, we introduce Iterative Markovian Fitting, a new methodology for solving SB problems, and Diffusion Schrödinger Bridge Matching (DSBM), a novel numerical algorithm for computing IMF iterates. DSBM significantly improves over previous SB numerics and recovers as special/limiting cases various recent transport methods. We demonstrate the performance of DSBM on a variety of problems.

Some statistical insights into PINNs

Claire Boyer

Sorbonne Université, Paris

Physics-informed neural networks (PINNs) combine the expressiveness of neural networks with the interpretability of physical modeling. Their good practical performance has been demonstrated both in the context of solving partial differential equations and in the context of hybrid modeling, which consists of combining an imperfect physical model with noisy observations. However, most of their theoretical properties remain to be established. We offer some food for thought and statistical insight into the proper use of PINNs.

Machine learning and optimal transport: some statistical and algorithmic tools

Elsa Cazelles

CNRS, Toulouse

In this talk, we focus the analysis of data that can be described by probability measures supported on a Euclidean space, by way of optimal transport. We aim at presenting a first and second order statistical analysis in the space of distributions in a concise manner, as a first approach to understand the general modes of variation of a set of observations. In the context of optimal transport, these studies correspond to the barycenter and the decomposition into geodesic principal components in the Wasserstein space. In particular, we consider a regularized estimator of the barycenter, in order to handle the noise coming from the observations.

Free Probability for predicting the performance fully connected feed-forward neural networks

Reda Chhaibi

Université Paul Sabatier, Toulouse

Joint work with Tariq DAOUDA & Ezéchiel KAHN. Based on `https://arxiv.org/abs/2111.00841`

Gradient descent during the learning process of a neural network can be subject to many instabilities. The spectral density of the Jacobian is a key component for analyzing stability. Following the works of Pennington et al., such Jacobians are modeled using free multiplicative convolutions from Free Probability Theory (FPT). We make the following contributions:

- theoretical: refine the metamodel of Pennington et al. thanks to the rectangular analogue of free multiplicative convolutions.
- numerical: present and benchmark a homotopy method for solving the equations of free probability.
- empirical: we show that the relevant FPT metrics computed before training are highly correlated to final test accuracies - up to 85%.

Counter-examples in first-order optimization: a constructive approach

Aymeric Dieuleveut

Ecole Polytechnique, Paris

This is a joint work with Baptiste Goujaud and Adrien Taylor.

While many approaches were developed for obtaining worst-case complexity bounds for first-order optimization methods in the last years, there remain theoretical gaps in cases where no such bound can be found. In such cases, it is often unclear whether no such bound exists (e.g., because the algorithm might fail to systematically converge) or simply if the current techniques do not allow finding them. In this work, we propose an approach to automate the search for cyclic trajectories generated by first-order methods. This provides a constructive approach to show that no appropriate complexity bound exists, thereby complementing the approaches providing sufficient conditions for convergence. Using this tool, we provide ranges of parameters for which some of the famous heavy-ball, Nesterov accelerated gradient, inexact gradient descent, and three-operator splitting algorithms fail to systematically converge, and show that it nicely complements existing tools searching for Lyapunov functions.

Polynomial Argmin for Recovery and Approximation of Multivariate Discontinuous Functions

Didier Henrion

CNRS, Toulouse

We propose to approximate a (possibly discontinuous) multivariate function $f(x)$ on a compact set by the partial minimizer $\arg\min_y p(x, y)$ of an appropriate polynomial p whose construction can be cast in a univariate sum of squares (SOS) framework, resulting in a highly structured convex semidefinite program. In a number of non-trivial cases (e.g. when f is a piecewise polynomial) we prove that the approximation is exact with a low-degree polynomial p. Our approach has three distinguishing features: (i) It is mesh-free and does not require the knowledge of the discontinuity locations. (ii) It is model-free in the sense that we only assume that the function to be approximated is available through samples (point evaluations). (iii) The size of the semidefinite program is independent of the ambient dimension and depends linearly on the number of samples. We also analyze the sample complexity of the approach, proving a generalization error bound in a probabilistic setting. This allows for a comparison with machine learning approaches. Joint work with Milan Korda and Jean Bernard Lasserre.

Two contributions to the theory of neural networks

François Malgouyres

Université Paul Sabatier, Toulouse

In this talk, I will first present results obtained by El-Mehdi Achour during his thesis, followed by results from Joachim Bona-Pellissier's thesis.

In this first part, we focus on the objective function landscape, for the quadratic cost function, of deep linear networks. I'll detail a weight-based characterization of critical point properties according to three categories: global minimizers, strict saddle points and non-strict saddle points. When the trajectory of ADAM iterates passes close to a critical point, I will show experimentally that the nature of the critical point has a very strong impact on the rest of the trajectory.

In the second part of the talk, for deep ReLU neural networks, I will study conditions linking the sample X and the parameters θ defining the network in relation to the local identifiability of θ. I will establish: 1/ A necessary and sufficient geometric condition for local identifiability. 2/ A computable sufficient condition and 3/ a computable necessary condition for local identifiability.

Optimistic Gradient Descent Ascent in General-Sum Bilinear Games

Jérôme Renault

Toulouse School of Economics

Joint with with Etienne de Montbrun (TSE-ANITI)

We study the convergence of Optimistic Gradient Descent Ascent (OGDA) in unconstrained bilinear games. For zero-sum games, we prove the exponential convergence of OGDA to a Nash equilibrium for any payoff matrix, and provide the exact ratio of convergence as a function of the step size. Then, we introduce OGDA for general-sum games, and show that in many cases, either OGDA converges exponentially fast to a Nash equilibrium, or the payoffs for both players converge to $+\infty$. We also show how to increase drastically the speed of convergence of a zero-sum problem, by introducing a general-sum game using the Moore-Penrose inverse of the original payoff matrix. This shows that general-sum games can be used to optimally improve algorithms designed for min-max problems.

FISTA is an automatic geometrically optimized algorithm for strongly convex functions.

Aude Rondepierre

Institut National des Sciences Appliquées, Toulouse

In this talk, we are interested in the famous FISTA algorithm. We show that FISTA is an automatic geometrically optimized algorithm for functions satisfying a quadratic growth assumption. This explains why FISTA works better than the standard Forward-Backward algorithm (FB) in such a case, although FISTA is known to have a polynomial asymptotic convergence rate while FB is exponential. We provide a simple rule to tune the α parameter within the FISTA algorithm to reach an ϵ-solution with an optimal number of iterations. These new results highlight the efficiency of FISTA algorithms, and they rely on new non asymptotic bounds for FISTA.

A Newton-type method for strict saddle functions

Clément Royer

Université Paris Dauphine-PSL

Nonconvex optimization problems arise in many fields of computational mathematics and data science. Recent advances in this area have focused on instances where nonconvexity has a benign effect, in the sense that traditional optimization algorithms exhibit good theoretical and practical performance on such problems. On the other hand, exploiting a particular nonconvex structure to improve algorithmic design is a more challenging endeavor.

In this talk, we describe a Newton-type framework for solving nonconvex optimization problems that exhibit a particular structure characterized by a strict saddle property. Our approach benefits from the local convergence properties of Newton's method, and is endowed with global convergence rates that improve over those known for generic nonconvex instances. In addition, our algorithm and its analysis can account for the presence of manifold constraints, which we illustrate on a selection of strict saddle functions.

Practical information

Workshop location:
ENSEEIHT, 2 rue Camichel, 31000 Toulouse.
Métro François Verdier.

Get there from the train station:
(time estimate by Google Maps, usually underestimated)

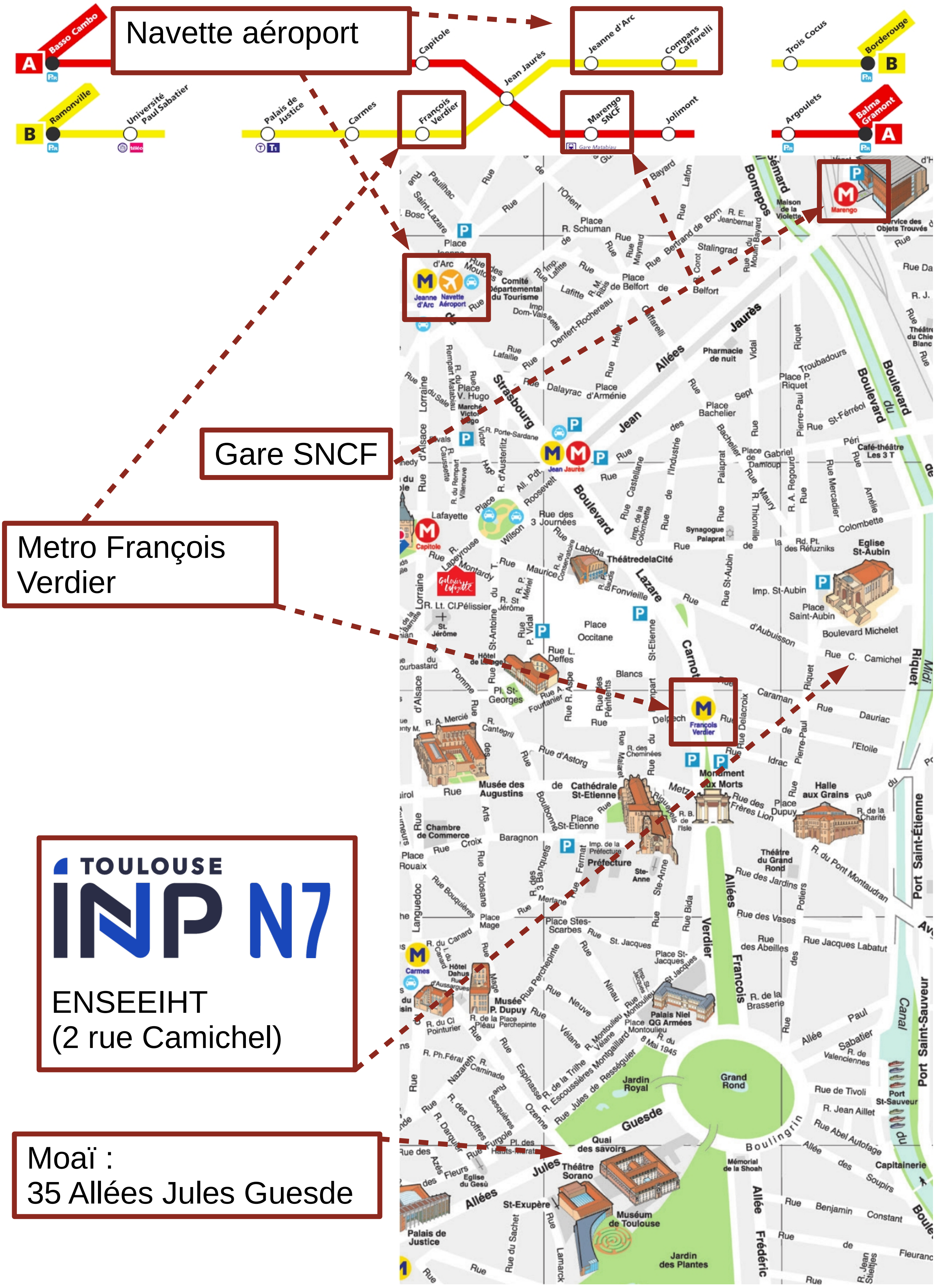
- Walk 17 minutes. Turn left when you exit the train station and follow the Canal du Midi.
- Bus 27: 3 minutes walk to Riquet stop (get in), Guilhemery stop (2 minutes, get out), cross the canal, 3 minutes walk, every 15 minutes.
- Bus L9: Marengo stop (get in), St-Georges stop (8 minutes, get out), 6 minutes walk, every 8 minutes.
- Metro A: Marengo-SNCF direction Basso-Cambo (get in), Jean Jaurès (1 stop, 1 minute, get out), 10 minute walk, every 3 minutes. May also commute at Jean Jaurès, take line B, direction Ramonville, get out at François Verdier (1 stop).
- Metro A+B: Commute at Jean Jaurès, line B direction Ramonville (get in), François Verdier (1 stop, 1 min, get out)), 7 minutes walk.

Get there from the airport:

- Bus "Navette Aéroport": at metro station Jeanne d'Arc or Compans Caffarelli (B line). 25 to 40 minutes (traffic jams from 8am to 10am and 4pm to 7pm). Every 15 minutes.
- Tram T2: the connection between downtown Toulouse and the aeroport is closed, we do not recommend using the tram to get to the airport.

Gala Dinner: Le Moaï, 35 allées Jules Guesde, Monday 26, 19:30 (registration mandatory).

- Walk 20 minutes from ENSEEIHT.
- Metro B: direction Ramonville, stop at Palais de Justice + 10 minutes walk.



Navette aéroport
Gare SNCF
Metro François Verdier
TOULOUSE INP N7
ENSEEIHT
(2 rue Camichel)
Moaï :
35 Allées Jules Guesde
A
Basso Cambo
Capitole
Jean Jaurès
Marengo SNCF
Jolimont
Argoulets
Balma Gramont
B
Ramonville
Université Paul Sabatier
Palais de Justice
Carmes
François Verdier
Jeanne d'Arc
Compans Caffarelli
Trois Cocus
Borderouge