Edinburgh · Scotland · May 24-25, 2012

Advances in Large-Scale Optimization

A NAIS Workshop, Trek and Colloquium

In modern digital world, with ever increasing amounts of readily-available data comes the need to solve optimization problems of unprecedented sizes. Machine learning, compressed sensing, natural language processing, truss topology design and computational genetics are some of many prominent application domains where it is easy to formulate optimization problems with millions of variables. Many modern optimization algorithms, while exhibiting great efficiency in modest dimensions, are not designed to scale to instances of this size and are hence often, unfortunately, not applicable. On the other hand, simple methods, some having been proposed decades ago, are experiencing a comeback — albeit in modern forms. This workshop aims to bring together researchers working on novel optimization algorithms capable of working in large-scale setting.

Organized by Peter Richtárik

NAIS Centre for Numerical Algorithms and Intelligent Software

THE UNIVERSITY OF EDINBURGH

MATHEMATICS FOR VAST DIGITAL RESOURCES

Headline Speaker

Yurii Nesterov

Yurii Nesterov is a professor at the Université catholique de Louvain, Louvain-la-Neuve, Belgium. He is affiliated with the Center for Operations Research and Econometrics (CORE) and the Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM).

His collaboration with Arkadi Nemirovski, resulting in the book Interior-Point Polynomial Algorithms for Convex Programming, has developed the theory of self-concordant functions to unify global complexity results obtained for convex optimization problems including linear, second-order cone and semi-definite programming. Many scholars consider this book to be the single most important contribution to optimization theory in the past twenty years.

More recently, Nesterov is also the author of the monograph Introductory Lectures on Convex Optimization, which develops state-of-the-art theory at a level appropriate for introductory graduate courses. In recent work he has obtained results on the global convergence of a regularized Newton's method for unconstrained optimization and established a theory of smoothing that allows for the applicability of optimal first-order methods to large-scale problems with nondifferentiable objectives. According to Google Scholar, the work of Prof. Nesterov has been cited more than 7,000 times.

Yurii Nesterov is the recipient of several awards and honors, including George B. Dantzig Prize (2000), the John Von Neumann Theory Prize (2009), and an invitation to address the International Congress of Mathematicians (2010). Dantzig Prize, awarded jointly by the Mathematical Optimization Society (MOS) and the Society for Industrial and Applied Mathematics (SIAM), is awarded for original research, which by its originality, breadth and scope, is having a major impact on the field of mathematical programming, once every three years. The John von Neumann Prize is awarded annually to a scholar who has made fundamental, sustained contributions to theory in operations research and the management sciences.

Workshop (May 24, 2012)

Venue

Informatics Forum
10 Crichton Street, George Square campus, University of Edinburgh (see attached map)

Schedule

09:15-09:35	*Tea and Coffee with Danish Selection*
09:35-09:45	*Welcome*
09:45-10:25	**Yurii Nesterov** (CORE, UCL, headline speaker) Subgradient methods for huge-scale optimization problems
10:25-11:00	*Break*
11:00-11:40	**Mark Schmidt** (INRIA) Inexact proximal-gradient methods and linearly-convergent stochastic gradient methods
11:45-12:25	**Peter Richtárik** (Edinburgh) Parallel block coordinate descent methods for huge-scale partially separable problems
12:25-14:00	*Lunch (Informatics Forum, 4th level, room: Mini Forum 2)*
14:00-14:40	**Alexandre D'Aspremont** (École Polytechnique) A stochastic smoothing algorithm for semidefinite programming
14:45-15:25	**Jacek Gondzio** (Edinburgh) Matrix-free interior point method for large-scale optimization
15:25-16:00	*Tea and Coffee with Biscuits*
16:00-16:40	**Michal Kočvara** (Birmingham) Domain decomposition techniques in topology optimization
16:45-17:25	**Coralia Cartis** (Edinburgh) A new and improved recovery analysis for iterative hard thresholding algorithms in compressed sensing
17:30-18:10	**Nathan Srebro** (Toyota Technological Institute at Chicago) Stochastic approximation with mini-batches: smoothness, optimistic rates and acceleration

Abstracts

Yurii Nesterov

Subgradient methods for huge-scale optimization problems

We consider a new class of huge-scale problems, the problems with sparse subgradients. The most important functions of this type are piece-wise linear. For optimization problems with uniform sparsity of corresponding linear operators, we suggest a very efficient implementation of subgradient iterations, the total cost of which depends logarithmically on the dimension. This technique is based on a recursive update of the results of matrix/vector products and the values of symmetric functions. It works well, for example, for matrices with few nonzero diagonals and for max-type functions.

We show that the updating technique can be efficiently coupled with the simplest subgradient methods, the unconstrained minimization method by B. Polyak, and the constrained minimization scheme by N. Shor. Similar results can be obtained for a new nonsmooth random variant of a coordinate descent scheme. We present also promising results of preliminary computational experiments.

Mark Schmidt

Inexact proximal-gradient methods and linearly-convergent stochastic gradient methods

The first part of this talk considers the problem of optimizing the sum of a smooth convex function and a non-smooth convex function using proximal-gradient methods, where an error is present in the calculation of the gradient of the smooth term or in the proximity operator with respect to the non-smooth term. We show that both the basic proximal-gradient method and the accelerated proximal-gradient method achieve the same convergence rate as in the error-free case, provided that the errors decrease at appropriate rates. Using these rates, we perform as well as or better than a carefully chosen fixed error level on a set of structured sparsity problems.

The second part of the talk considers optimizing the sum of a large number of smooth functions, where the sum is strongly convex. Stochastic gradient algorithms only compute a single term in the sum on each iteration, leading to inexpensive iterations but unfortunately
yielding sublinear convergence rates. In contrast, full-gradient methods achieve linear convergence rates at the expense of evaluating all terms on each iteration. We explore two strategies that exhibit the benefits of both approaches. First, we show that a linear convergence rate can be achieved at the expense of evaluating an increasing number of functions on each iteration. Second, and more surprisingly, we propose a new method that only needs to evaluate a single term in the sum on each iteration but that still achieves a linear convergence rate. Numerical experiments indicate that the new algorithms can dramatically outperform standard methods.

Peter Richtárik

Parallel block coordinate descent methods for huge-scale partially separable problems

In this work we show that randomized block coordinate descent methods can be accelerated by parallelization when applied to the problem of minimizing the sum of a partially block separable smooth convex function and a simple block separable convex function. We give a generic algorithm and several variants thereof based on the way parallel computation is performed. In all cases we prove iteration complexity results, i.e., we give bounds on the number of iterations sufficient to approximately solve the problem with high probability. Our results generalize the intuitive observation that in the separable case the theoretical speedup caused by parallelization must be equal to the number of processors. We show that the speedup increases with the number of processors and with the degree of partial separability of the smooth component of the objective function. Our analysis also works in the mode when the number of blocks being updated at each iteration is random, which allows for modelling situations with variable (busy or unreliable) number of processors. We conclude with some encouraging computational results applied to huge-scale LASSO instances.

Joint work with Martin Takáč (University of Edinburgh).

Alexandre D'Aspremont

A stochastic smoothing algorithm for semidefinite programming

We use low rank stochastic perturbation techniques to produce smooth approximations to the maximum eigenvalue function. We show in particular that only three maximum eigenvector computations are sufficient to produce a Lipschitz continuous approximation to this function. We then use this result to design efficient semidefinite programming solvers.

Jacek Gondzio

Matrix-free interior point method for large-scale optimization

A redesign of interior point methods will be addressed. It has two objectives: (a) to avoid an explicit access to the problem data and to allow only matrix-vector products to be executed with the Hessian and Jacobian and its transpose; and (b) to allow the method work in a limited-memory regime.

Michal Kočvara

Domain decomposition techniques in topology optimization

We investigate several approaches to the solution of topology optimization problems using decomposition of the computational domain. We will focus on decomposing the whole (nonlinear constrained) optimization problem, rather than just on solving linear equations arising in the Newton method by standard domain decomposition methods.

Joint work with D. Loghin and J. Turner (both University of Birmingham).

Coralia Cartis

A new and improved recovery analysis for iterative hard thresholding algorithms in compressed sensing

We present a novel average-case analysis for a standard compressed sensing algorithm, Iterative Hard Thresholding (IHT). By analysing the fixed points of IHT, we obtain a condition for general measurement matrices guaranteeing at most one fixed point, namely the original signal. We also give an improved requirement guaranteeing the algorithm's convergence to some fixed point. Thus, if both conditions are satisfied, we ensure recovery of the original signal. For the specific case of Gaussian measurement matrices and independent signals, comparison with existing worst-case results by means of the phase transition framework shows a substantial quantitative improvement. Our analysis also applies to IHT variants with variable stepsize, such as normalized IHT, yielding again a significant improvement for the phase transition of Gaussian measurement matrices.

Joint work with Andrew Thompson (University of Edinburgh).

Nathan Srebro

Stochastic approximation with mini-batches: smoothness, optimistic rates and acceleration

Stochastic Approximation (SA) methods, such as Stochastic Gradient Descent, are often the methods of choice in large-scale learning problems. The typical approach is to use a single training example per iteration, in order to obtain a very rough gradient estimate. I will discuss how, for smooth objectives, use of multiple training examples per iteration can be justified, especially when parallel processing is considered, but in order to obtain parallel speeds in some regimes, acceleration is necessary. To this end, I will present a new variant of the accelerated gradient method which allows for stochasticity in the gradient estimates, and achieves good rates on the relative error for non-negative objectives (so-called "optimistic rates'').

Joint work with Andy Cotter, Ohad Shamir and Karthik Sridharan.

Workshop Participants

Iskander Aliev	Cardiff U, Mathematics
Yusuf Aytar	University of Oxford, Engineering Science
Bubacarr Bah	U of Edinburgh, Mathematics
Alexander Banks-Watson	U of Edinburgh, Mathematics
Euan Barlow	U of Strathclyde, Engineering
Burak Buke	U of Edinburgh, Mathematics
Waqquas Bukhsh	U of Edinburgh, Mathematics
Coralia Cartis	U of Edinburgh, Mathematics
Jessica Chen-Burger	U of Edinburgh, Informatics
Alexander D'Aspremont	Ecole Polytechnique
Paolo Favaro	Heriot-Watt U, Engineering
Bartosz Filipecki	U of Edinburgh, Mathematics
Roger Fletcher	U of Dundee, Mathematics
Michelle Galea	U of Edinburgh, Informatics
Cong Geng	Nanyang Technological U
Jacek Gondzio	U of Edinburgh, Mathematics
Pablo Gonzalez-Brevis	U of Edinburgh, Mathematics
Ivan Grebenkin	Heriot-Watt U, Engineering
Andreas Grothey	U of Edinburgh, Mathematics
Cristóbal Guzmán	Georgia Institute of Technology, Engineering
Morteza Haghighat Sefat	Heriot-Watt U, Engineering
Julian Hall	U of Edinburgh, Mathematics
Sophie Harland	U of Edinburgh, GeoSciences
Shaun Kelly	U of Edinburgh, Engineering
Michal Kočvara	U of Birmingham, Mathematics
Anna Kononova	Heriot-Watt U, Mathematics and CS
Ľubor Ladický	University of Oxford, Engineering Science
Jan Lellmann	U of Cambridge, Mathematics
Martin Lotz	U of Edinburgh, Mathematics
Robert Luce	TU Berlin, Mathematics
Duy V.N. Luong	Imperial College London, Computing
Hassan Mahmud	U of Edinburgh, Informatics
Jakub Mareček	U of Edinburgh, Mathematics
Ken McKinnon	U of Edinburgh, Mathematics
Mojtaba Moradi	Heriot-Watt U, Engineering
Sean Moran	U of Edinburgh, Informatics
Khafiz Muradov	Heriot-Watt U, Engineering
Yurii Nesterov	U catholique de Louvain, CORE
Tri-Dung Nguyen	U of Southampton, Mathematics & Management
Peter Orchard	U of Edinburgh, Informatics

Iliana Peneva	U of Edinburgh, Mathematics
Houduo Qi	U of Southampton, Mathematics
Peter Richtárik	U of Edinburgh, Mathematics
Michael Sapienza	Oxford Brookes U, Computing
Dmitry Savostyanov	Russian Academy of Sciences, Mathematics
Mark Schmidt	INRIA, Informatics
Tim Schulze	U of Edinburgh, Mathematics
Sunando Sengupta	Oxford Brookes U, Computing
Athina Spiliopoulou	U of Edinburgh, Informatics
Nathan Srebro	Toyota Technological Institute at Chicago
Paul Sturgess	Oxford Brookes U, Computing
Martin Takáč	U of Edinburgh, Mathematics
Jared Tanner	U of Edinburgh, Mathematics
Rachael Tappenden	U of Edinburgh, Mathematics
Vibhav Vineet	Oxford Brookes U, Computing
Kristian Woodsend	U of Edinburgh, Mathematics
Mehrdad Yaghoobi	U of Edinburgh, Engineering
Yiming Yan	U of Edinburgh, Mathematics
Faraj Zarej	Heriot-Watt U, Engineering
Ziming Zhang	Oxford Brookes U, Computing
Shuai Zheng	Oxford Brookes U, Computing

Trek to Arthur's Seat (May 25, 2012)

On Friday morning there is an optional walk / light trek to the top of Arthur's Seat—a 350 million year old volcano in the city centre rising 251 meters above the sea level—offering magnificient views of Edinburgh and the Firth of Forth. We are departing at 9:30 sharp from the entrance of Informatics Forum, please plan to arrive 10 minutes earlier.

"The views from the summit [of Arthur's seat] are awesome. The city itself is a real gem, Edinburgh is in my own opinion the most beautiful and fascinating city in Britain by a mile." www.TrekkingBritain.com

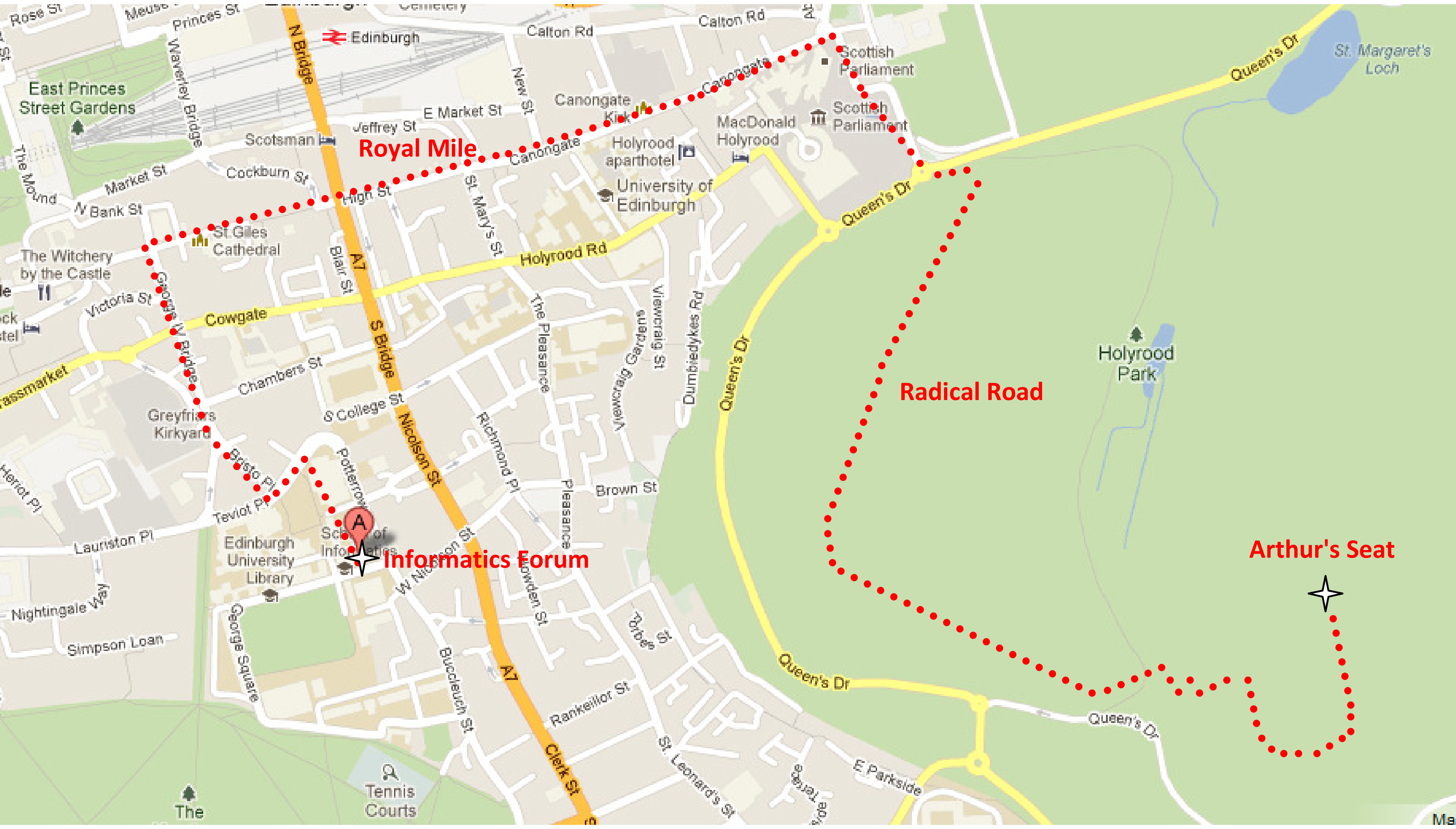
While the walk is reasonably light, good footwear is recommended as there will be some ascending to do. It is recommended that you bring along a light weather-proof jacket; it can get windy up on the hill. Also, please bring enough water and light refreshments with you.

Schedule

09:20-09:30	*Meeting in front of the Informatics Forum*
09:30-12:30	Light trek to the top of Arthur's Seat (and back to Royal Mile)
12:30-14:00	*Lunch (individual)*

The Route

We are not going to walk straight to the Arthur's Seat. Instead, we will take a detour and first walk through the George Square campus towards the National Museum of Scotland (entrance free, great exhibits) and the Elephant house, the birthplace cafe of Harry Potter. Continuing along the 300m long George IV Bridge built in 1832, we enter the Royal Mile (a 1 mile long backbone of Edinburgh's historic centre joining the Edinburgh Castle and the Holyrood Abbey) near the High Court of Justiciary, Scotland's supreme criminal court. From that point we will walk along the Royal Mile, away from the Castle and towards the Holyrood Abbey, i.e., down the hill, passing many points of interest which you might want to visit if your are staying in Edinburgh a bit longer: St. Gile's Cathedral (Mother Church of Presbyterianism), Mary King's Close, North Bridge, Canongate Kirk, Scottish Parliament Building, Dynamic Earth Science Centre, Palace of Holyroodhouse and Queen's Gallery. Once at the end of the Royal Mile we enter the Holyrood Park, walking along a foot path (called Radical Road) below Salisbury Crags—a stretch of over 46 meters high cliffs of dolerite and columnal basalt offering spectacular views of the city centre. At the end of the path is the foot of Arthur's Seat; from there it is just a 20 min walk up to the summit point. We will return to the Royal Mile via a different route through the Holyrood Park, passing by the St. Margaret's Loch and the ruins of St Anthony's Chapel.



Colloquium (May 25, 2012)

Venue

The Colloquium is held in the James Clerk Maxwell Building (JCMB) of the King's Buildings campus of the University of Edinburgh, in the southern part of the city. After a coffee and tea in 5th floor common room (the room on the left when getting off the elevator) at 15:30, the lecture starts in Lecture Theatre C at 16:00 (on the third floor of JCMB). Following the lecture there will be a Wine and Cheese Reception in the Costa Coffee area (third floor of JCMB) at 17:00.

Schedule

15:30-16:00	*Tea and Coffee with Carrot Cake (JCMB 5th floor, room 5212)*
16:00-17:00	**Yurii Nesterov** (CORE, UCL, headline speaker) Optimization in relative scale
17:00-19:00	*Wine & Cheese Reception (JCMB 3rd floor; Costa Coffee area)*

Abstract

Yurii Nesterov

Optimization in relative scale

The majority of modern optimization methods can deliver approximate solutions with absolute accuracy. Usually, the quality of such an approximations is affected by unknown parameters of the problems like Lipschitz constants and the size of the optimal solution. As a consequence, very often we ask for much higher absolute accuracy than needed.

In the past few years we see more and more examples of optimization models where the optimal solution is positive by its construction. For such problems it is usually possible to develop special schemes which work with relative accuracy. Usually, such schemes need a preprocessing phase for computing an appropriate metric for measuring the problem data. Complexity estimates for such problems often depend only on the dimension of the problem and the desired relative accuracy. In our talk, we will discuss the most important examples of this approach.

Maps

The University in the city

The University Central Area

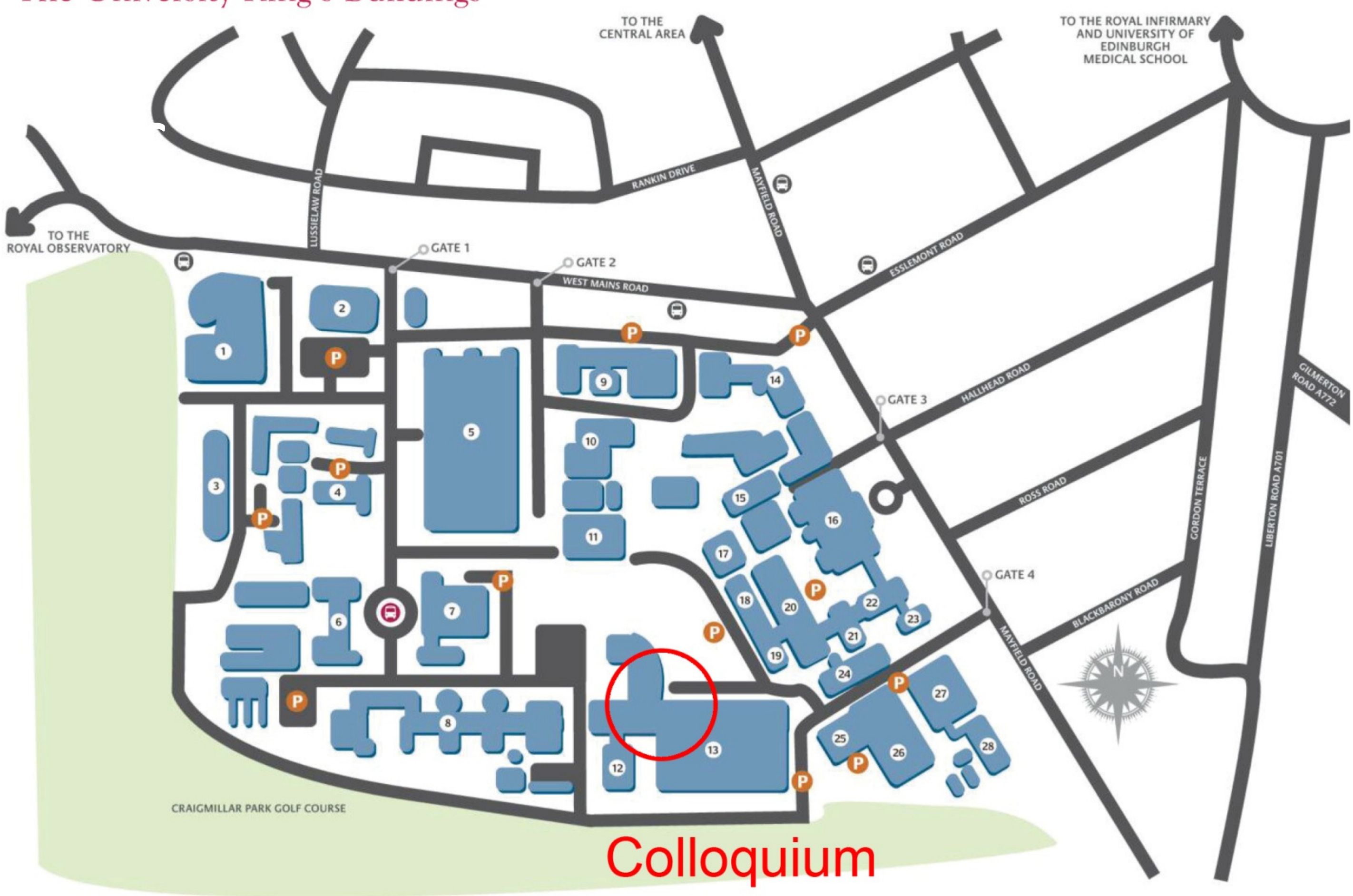
1 New College: School of Divinity
2 Minto House
3 Charles Stewart House
4 Adam House
5 St Cecilia's Hall
6 Old Infirmary (Geography)
7 Old Surgeon's Hall
8 Drummond Library (GeoSciences)
9 Chisholm House
10 The Moray House School of Education
11 St Leonard's Land
12 Centre for Sport and Exercise
13 The Pleasance
14 Old College
15 Alison House
16 The Potterrow Student Centre: Students' Association (EUSA); Chaplaincy Centre; Health Centre
17 Bedlam Theatre
18 McEwan Hall
19 Medical School
20 Chrystal Macmillan Building
21 Hugh Robson Building
22 Reid Concert Hall
23 7 George Square
24 Teviot Row House: Student Union
25 1 George Square
26 Dugald Stewart Building
27 Visitor Centre: Information, Exhibition and Shop
28 Informatics Forum
29 Appleton Tower
30 International Office and Student Recruitment and Admissions
31 College of Humanities & Social Science Office
32 50 George Square
33 David Hume Tower Lecture Theatres
34 David Hume Tower
35 Business School
36 George Square Lecture Theatre
37 Main Library; Careers Service; Student Counselling Service; Student Disability Service
38 International Student Centre
39 Office of Lifelong Learning
40 Institute for Advanced Studies in the Humanities (IASH)
Shuttle bus to King's Buildings Campus
41 Edinburgh Central Mosque
42 Surgeon's Hall
43 Royal Scottish Museum
44 Greyfriars Kirk
45 The National Library of Scotland
46 St Giles' Cathedral
47 Tron Kirk
48 City Chambers
49 Edinburgh Castle
50 Royal Scottish Academy
51 National Gallery of Scotland
52 Edinburgh Waverley train station
53 St Andrew's bus station
Public bus stop

NAIS Centre for Numerical Algorithms and Intelligent Software

The University of Edinburgh

Mathematics for Vast Digital Resources

The University King's Buildings



The regular bus service between the Central area and King's Buildings can be viewed at
www.ed.ac.uk/schools-departments/transport/public-transport/buses/shuttle-bus

1 British Geological Survey
2 Student residences
3 Weir Building: Science & Engineering College Office; Careers Service; Student Counselling Service
4 Crew Building
5 Joseph Black Building
6 Scottish Micro Electronics Centre
7 Roger Land Building
8 Peter Wilson Building
9 Grant Institute
10 KB House: Students' Association (EUSA)
11 KB Centre
12 Erskine Williamson Building
13 James Clerk Maxwell Building
14 Ashworth Laboratories
15 Kenneth Denbigh Building
16 Sanderson Building
17 John Muir Building
18 William Rankine Building
19 Alexander Graham Bell Building
20 Fleeming Jenkin Building
21 Faraday Building
22 Hudson Beare Building
23 Engineering Lecture Theatre
24 Alrick Building
25 Michael Swann Building
26 Darwin Building
27 Daniel Rutherford Building
28 CH Waddington Building

Shuttle bus to Central Area Campus
Public bus stop
Parking

Internet Access

eduroam

Please use eduroam if you have signed up for it at your home institution.

The eduroam network is a secure world-wide roaming access service for the international research and education community. Eduroam is available across Europe, North America and Asia/Pacific. If you are registered for eduroam at your home institution, then you should be able to connect automatically to the wireless network here. If this does not work, it may be because your computer has already cached site-specific details for a different eduroam location, and these conflict with the settings here. Try deleting your eduroam connection entirely and letting it bootstrap to pick up the local connection details.

- Network name (SSID): **eduroam**
- Wireless Network Authentication type: **WPA2 Enterprise**
- Data Encryption method: **AES**

If your wireless device does not support WPA2/AES, select WPA authentication and TKIP encryption.

Individual Wireless Accounts

Individual wireless access accounts will be provided to those who are not able to use eduroam.

Please ask at the registration desk or talk to Rachael Tappenden / Kristian Woodsend if you have not received one.

Acknowledgements

Thanks to Rachael Tappenden, Jakub Mareček , Martin Takáč, Michelle Swain and Magdalena Mazurczcak for helping with various organizational details.

Notes

Notes

Notes