What's New in Gurobi 9.1

GUROBI OPTIMIZATION

The World's Fastest Solver

Tobias Achterberg
08/09 December 2020

Gurobi 9.1 – Summary

- Performance improvements compared to Gurobi 9.0
  - See separate talk on performance improvements in Gurobi 9.1
- PSD Cuts for non-convex MIQCPs
  - See separate talk on non-convex MIQCP enhancements
- IntegralityFocus parameter
  - Avoid numerical issues with trickle flows
- Heuristics
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- Pre-specified user cuts
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- Multi-objective
  - Better control of objective relaxation constant
- IIS size estimate
  - See estimate of size of IIS in log file
- Python enhancements
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- Compute Server communication protocol performance
  - Reduced number of messages passed between client and server
- Record/replay for Compute Server and Cloud
- Tuning tool enhancements
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- Additional Parameter
  - PoolGapAbs parameter
- Additional Attribute
  - FingerPrint attribute

Gurobi 9.1 – Summary

- Performance improvements compared to Gurobi 9.0
  - See separate talk on performance improvements in Gurobi 9.1
- PSD Cuts for non-convex MIQCPs
  - See separate talk on non-convex MIQCP enhancements
- IntegralityFocus parameter
  - Avoid numerical issues with trickle flows
- Heuristics
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- Pre-specified user cuts
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- Multi-objective
  - Better control of objective relaxation constant
- IIS size estimate
  - See estimate of size of IIS in log file
- Python enhancements
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- Compute Server communication protocol performance
  - Reduced number of messages passed between client and server
- Record/replay for Compute Server and Cloud
- Tuning tool enhancements
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- Additional Parameter
  - PoolGapAbs parameter
- Additional Attribute
  - FingerPrint attribute

Integrality Focus – Trickle Flows

- **Very common model construct:**

$$z = 0 \rightarrow x = 0, 0 \leq x \leq M, z \text{ binary}$$

  - If you don't build a factory, you cannot produce anything
- **Modeling options**
  - Indicator constraint $z = 0 \rightarrow x \leq 0$
    - Automatically translated into SOS
  - SOS constraint $z + z' = 1, SOS1(z', x)$
    - Depending on "PreSOS1BigM", automatically translated into Big-M
  - Big-M constraint $x \leq Mz$
- **Numerical precision vs. performance**
  - SOS is not part of LP relaxation: bad performance
  - Big-M constraint allows trickle flow: $z = 10^{-5}, x = 10^{-5}M$

Integrality Focus – Trickle Flows

GUROBI OPTIMIZATION

- **Our solution (IntegralityFocus = 1)**
  - Round integers and check feasibility
  - If not feasible, reject solution and branch on integer
  - Domain propagation in child node may resolve trickle flow
- **Example**

$$x - 10^8 z_1 - 10^8 z_2 \leq 0$$
$$x^* = 1100, z_1^* = 10^{-6}, z_2^* = 10^{-5}$$

  - Round integers to zero: constraint is violated by 1100
  - Branch on $z_1 \leq 0$
    - New solution: $x^* = 1100, z_1^* = 10^{-6}, z_2^* = 10^{-5}$
    - Bad luck: current solution is still feasible in tolerance, hence no simplex pivot applied
  - Continue in $z_1 \leq 0$ child node by branching on $z_2 \leq 0$
    - Now, domain propagation concludes $x \leq 0$
    - New solution: $x^* = 0, z_1^* = 10^{-6}, z_2^* = 10^{-6}$
    - Accept solution as feasible

Integrality Focus – Trickle Flows

- Impact on numerical stability

Trickle Flow Viol	>10.0	>1.0	$>10^{-1}$	$>10^{-2}$	$>10^{-3}$	$>10^{-4}$	$>10^{-5}$
Default	5	15	7	7	22	14	28
IntegralityFocus = 1	0	0	0	0	2	7	8

- Impact on performance
  - 255 models out of ~4000 affected
  - 20% slower on affected models on average
  - 2% slower overall
- Not a default setting
  - User needs to activate it explicitly
  - Similar to NumericFocus:
    - It may hurt performance
    - It may help with numerics, but is not guaranteed to do so

Gurobi 9.1 – Summary

GUROBI OPTIMIZATION

- Performance improvements compared to Gurobi 9.0
  - See separate talk on performance improvements in Gurobi 9.1
- PSD Cuts for non-convex MIQCPs
  - See separate talk on non-convex MIQCP enhancements
- IntegralityFocus parameter
  - Avoid numerical issues with trickle flows
- Heuristics
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- Pre-specified user cuts
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- Multi-objective
  - Better control of objective relaxation constant
- IIS size estimate
  - See estimate of size of IIS in log file
- Python enhancements
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- Compute Server communication protocol performance
  - Reduced number of messages passed between client and server
- Record/replay for Compute Server and Cloud
- Tuning tool enhancements
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- Additional Parameter
  - PoolGapAbs parameter
- Additional Attribute
  - FingerPrint attribute

NoRel Heuristic

- **Primary target: models for which LP relaxation is hard to solve**
  - LP takes too long, but user wants to have a good feasible solution in reasonable amount of time
  - Tough for our other heuristics
    - Sophisticated ones require an LP solution to start from
    - Existing Non-LP heuristics are very cheap, usually produce low-quality solutions, if any
- **Secondary target: find best solution in given time limit**
  - Sometimes, node throughput is too small
    - Not enough exploration to find good solutions in given time limit
  - NoRel heuristic searches in a more diverse solution space
    - May find more and better solutions than regular branch-and-cut
  - Downside: no objective bound and solution quality assessment

NoRel Heuristic

- **NoRel heuristic is called between presolve and initial root LP**
  - Start from some (feasible or infeasible) vector
    - Constructed by quick heuristic
  - Solve smaller sub-MIPs to decrease infeasibility or objective value
    - Use multiple threads to solve sub-MIPs in parallel
  - Various neighborhood strategies
    - adaptive to spend more time on more successful ones
  - Basically runs forever until specified time or work limit is hit
- **Parameters to control NoRel heuristic**
  - NoRelHeurTime
    - Specifies time to spend in heuristic
    - non-deterministic, based on wall-clock time
  - NoRelHeurWork
    - Specifies work to spend in heuristic
    - deterministic

Gurobi 9.1 – Summary

- Performance improvements compared to Gurobi 9.0
  - See separate talk on performance improvements in Gurobi 9.1
- PSD Cuts for non-convex MIQCPs
  - See separate talk on non-convex MIQCP enhancements
- IntegralityFocus parameter
  - Avoid numerical issues with trickle flows
- Heuristics
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- Pre-specified user cuts
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- Multi-objective
  - Better control of objective relaxation constant
- IIS size estimate
  - See estimate of size of IIS in log file
- Python enhancements
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- Compute Server communication protocol performance
  - Reduced number of messages passed between client and server
- Record/replay for Compute Server and Cloud
- Tuning tool enhancements
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- Additional Parameter
  - PoolGapAbs parameter
- Additional Attribute
  - FingerPrint attribute

Pre-Specified User Cuts

- **Up to Gurobi 9.0: user cuts could only be added via callback**
- **Gurobi 9.1: user cuts can also be provided explicitly in advance**
  - Similar to lazy constraints
    - Add as regular linear constraint
    - Set the “Lazy” attribute to -1
- **Recall properties of user cuts:**
  - Must be redundant w.r.t. other constraints and integrality
    - Hence, without them, set of feasible solutions stays identical
  - Will be considered to be added to LP relaxation if they cut off current LP solution
    - Apply same filtering mechanism as for Gurobi’s internal cuts

Gurobi 9.1 – Summary

- Performance improvements compared to Gurobi 9.0
  - See separate talk on performance improvements in Gurobi 9.1
- PSD Cuts for non-convex MIQCPs
  - See separate talk on non-convex MIQCP enhancements
- IntegralityFocus parameter
  - Avoid numerical issues with trickle flows
- Heuristics
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- Pre-specified user cuts
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- Multi-objective
  - Better control of objective relaxation constant
- IIS size estimate
  - See estimate of size of IIS in log file
- Python enhancements
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- Compute Server communication protocol performance
  - Reduced number of messages passed between client and server
- Record/replay for Compute Server and Cloud
- Tuning tool enhancements
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- Additional Parameter
  - PoolGapAbs parameter
- Additional Attribute
  - FingerPrint attribute

IIS Size Estimate

Constraints Min	Constraints Max	Constraints Guess	Bounds Min	Bounds Max	Bounds Guess	Runtime
0	4055	-	0	5476	-	0s
0	4038	-	0	5476	-	35s
0	4038	460	0	5476	-	40s
0	4032	460	0	5476	-	45s
0	4032	380	0	5476	-	51s
0	4014	350	0	5476	-	56s
0	4008	360	0	5476	-	75s
0	4003	360	0	5476	-	80s
0	3998	440	0	5476	-	87s
1	3991	490	0	5476	-	115s
1	3990	490	0	5476	-	121s
2	3988	490	0	5476	-	126s
3	3987	510	0	5476	-	130s
4	3984	610	0	5476	-	136s
4	3982	550	0	5476	-	140s
6	3980	690	0	5476	-	146s

- **Often pretty good estimate of the size of the final IIS**
  - Just an estimate; neither an upper or lower bound

Gurobi 9.1 – Summary

- **Performance improvements compared to Gurobi 9.0**
  - See separate talk on performance improvements in Gurobi 9.1
- **PSD Cuts for non-convex MIQCPs**
  - See separate talk on non-convex MIQCP enhancements
- **IntegralityFocus parameter**
  - Avoid numerical issues with trickle flows
- **Heuristics**
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- **Pre-specified user cuts**
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- **Multi-objective**
  - Better control of objective relaxation constant
- **IIS size estimate**
  - See estimate of size of IIS in log file
- **Python enhancements**
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- **Compute Server communication protocol performance**
  - Reduced number of messages passed between client and server
- **Record/replay for Compute Server and Cloud**
- **Tuning tool enhancements**
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- **Additional Parameter**
  - PoolGapAbs parameter
- **Additional Attribute**
  - FingerPrint attribute

Enhancements in gurobipy Python API

- **New `MConstr` class for matrix constraints**
  - Single object to model an array-like set of linear constraints
  - Example:

```
mc = model.addConstr(A @ x == b)
```

  - Attribute access via NumPy's ndarray:

```
mc.b = np.ones(n)  # Change RHS to all-ones
duals = mc.Pi  # Query dual values for these linear constraints
```

  - Enables concise, natural expressions:

```
print(norm(mc.rhs - A @ x.X))  # primal residual norm for Ax=b
print(abs(x.obj @ x.X - mc.rhs @ mc.pi))  # duality gap
```

Enhancements in gurobipy Python API

- Install gurobipy via pip:

```
pip install -i https://pypi.gurobi.com gurobipy
```

- Package hosted at Gurobi
  - Available on PyPI.org some time later
- Installation is self contained
  - No need to install the full Gurobi software distribution!
- Comes with limited-size trial license
  - Runs out of the box for limited-size models
  - Easy to plug in full (academic or commercial) license
- Supported Python versions:
  - 2.7 (deprecated)
  - 3.6, 3.7, 3.8, 3.9

Gurobi 9.1 – Summary

- Performance improvements compared to Gurobi 9.0
  - See separate talk on performance improvements in Gurobi 9.1
- PSD Cuts for non-convex MIQCPs
  - See separate talk on non-convex MIQCP enhancements
- IntegralityFocus parameter
  - Avoid numerical issues with trickle flows
- Heuristics
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- Pre-specified user cuts
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- Multi-objective
  - Better control of objective relaxation constant
- IIS size estimate
  - See estimate of size of IIS in log file
- Python enhancements
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- Compute Server communication protocol performance
  - Reduced number of messages passed between client and server
- Record/replay for Compute Server and Cloud
- Tuning tool enhancements
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- Additional Parameter
  - PoolGapAbs parameter
- Additional Attribute
  - FingerPrint attribute

Record/Replay for Compute Server and Gurobi Instant Cloud

- Use as in the local case
  - Just set the “Record” parameter to 1 and run your Compute Server or Gurobi Instant Cloud code
- Privacy
  - Will not store some string parameter settings in recording file
    - User credentials and secrets
    - File names (like log file names, nodefile directories)
    - Compute Server URL
  - Only stores that the parameter has been changed, but not to which value
  - In replay, use pre-defined names for these string parameters
  - To make it work, may need to set OS environment variables to inject string parameter values that work in the replay environment
    - e.g., a different Compute Server URL

Gurobi 9.1 – Summary

- Performance improvements compared to Gurobi 9.0
  - See separate talk on performance improvements in Gurobi 9.1
- PSD Cuts for non-convex MIQCPs
  - See separate talk on non-convex MIQCP enhancements
- IntegralityFocus parameter
  - Avoid numerical issues with trickle flows
- Heuristics
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- Pre-specified user cuts
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- Multi-objective
  - Better control of objective relaxation constant
- IIS size estimate
  - See estimate of size of IIS in log file
- Python enhancements
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- Compute Server communication protocol performance
  - Reduced number of messages passed between client and server
- Record/replay for Compute Server and Cloud
- Tuning tool enhancements
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- Additional Parameter
  - PoolGapAbs parameter
- Additional Attribute
  - FingerPrint attribute

Tuning Tool

- **New command line option TuneBaseSettings**
  - Example: grbtune TuneBaseSettings="base1.prm base2.prm base3.prm"
  - Tries these parameters first, in this order
    - As usual, subsequent parameter runs are aborted if they hit runtime of best parameter setting found so far
  - An empty parameter file corresponds to default settings
  - If the list does not contain an empty parameter file, then default settings will be appended to the list
    - Hence, in this case defaults are after all specified parameter settings
  - Can help to avoid long default run if user knows good parameter settings

Tuning Tool

- **New command line option TuneCleanup**
  - Parameter in [0,1]
  - Specifies that the last x% of runtime should be used for cleanup
  - Goal of cleanup: reduce the number of non-default parameter settings to get good performance
- **Additional output**
  - Display more detailed summary at end of tuning

```
Tested 17 parameter sets in 84.92s

Baseline parameter set: mean runtime 2.90s

        Default parameters

 # Name              0        1        2      Avg  Std Dev
 0 Model         3.12s    2.62s    2.96s    2.90s     0.21

Improved parameter set 1 (mean runtime 2.14s):

        Cuts 3

 # Name              0        1        2      Avg  Std Dev
 0 Model         2.11s    2.47s    1.86s    2.14s     0.25
```

Gurobi 9.1 – Summary

- **Performance improvements compared to Gurobi 9.0**
  - See separate talk on performance improvements in Gurobi 9.1
- **PSD Cuts for non-convex MIQCPs**
  - See separate talk on non-convex MIQCP enhancements
- **IntegralityFocus parameter**
  - Avoid numerical issues with trickle flows
- **Heuristics**
  - NoRel heuristic to find solutions if LP relaxation is too hard to solve
  - New variant of RINS improves ability to find good solutions early on
  - New heuristic for models with cardinality constraints
- **Pre-specified user cuts**
  - Provide explicit user cuts by adding linear constraints with Lazy=-1 attribute
- **Multi-objective**
  - Better control of objective relaxation constant
- **IIS size estimate**
  - See estimate of size of IIS in log file
- **Python enhancements**
  - Python matrix API enhancements
  - Pip install support
  - Release GIL during optimization
- **Compute Server communication protocol performance**
  - Reduced number of messages passed between client and server
- **Record/replay for Compute Server and Cloud**
- **Tuning tool enhancements**
  - TuneBaseSettings command line argument
  - TuneCleanup parameter
- **Additional Parameter**
  - PoolGapAbs parameter
- **Additional Attribute**
  - FingerPrint attribute