HiOp – User Guide

version 0.1

by

Cosmin G. Petra

Center for Applied Scientific Computing
Lawrence Livermore National Laboratory

7000 East Avenue,
Livermore, CA 94550, USA.

Oct 15, 2017
Updated Nov 03, 2019

Contents

1 Introduction

This document describes `HiOp`, a memory-distributed high-performance computing (HPC) quasi-Newton solver for large-scale nonlinear programming problems (NLPs) with a small number of general constraints.

This document includes instructions on how to obtain and build `HiOp` and a description of its interface, user options, and use as an optimization library. Guidelines on how is best to use the solver for parallel computations are also provided. The document generally targets users of `HiOp`, but also contains information relevant to potential developers or advanced users; these are strongly encouraged to also read the paper on the computational approach implemented in `HiOp` [2].

While `Hiop` solver targets DAE- and PDE-constrained optimization problems formulated in a "reduced-space" approach, it can be used for general nonconvex nonlinear optimization as well. For efficiency considerations, it is recommended to *use Hiop for NLPs that have a relatively small number of general constraints*, say less than 100; note that there are no restrictions on the number of bounds constraints, *e.g.*, one can specify simple bounds on any, and potentially all the decision variables without affecting the computational efficiency. The minimizers computed by `HiOp` satisfies *local* first-order optimality conditions.

The goal of `HiOp` is to remove the parallelization limitations of existing state-of-the-art solvers for nonlinear programming (NLP) and match/surpass the parallel scalability of the underlying PDE or DAE solver. Such limitation occurs whenever the dimensionality of the optimization space is as large as the dimensionality of the discretization of the differential systems of equations governing the optimization. In these cases, the use of existing NLP solvers results in i. considerable long time spent in optimization, which affects the parallel scalability, and/or ii. memory requirements beyond the memory capacity of the computational node that runs the optimization. `HiOp` removes these scalability/parallelization bottlenecks (for certain optimization problems described above) by offering interface for a *memory-distributed* specification of the problem and parallelizing the optimization search using specialized parallel linear algebra technique.

The general computational approach in `Hiop` is to use existing state-of-the-art NLP algorithms and develop linear algebra kernels tailored to the specific of this class of problems. `HiOp` is based on an interior-point line search filter method [3, 4] and follows the implementation details from [5], which is the implementation paper for IPOPT open-source NLP solver. The quasi-Newton approach is based on limited-memory secant approximations of the Hessian [1], which is generalized as required by the specific of interior-point methods for constrained optimization problems [2]. The specialized linear algebra decomposition is obtained by using a Schur-complement reduction that leverages the fact that the quasi-Newton Hessian matrix has a small number of dense blocks that border a low-rank update of a diagonal matrix. The technique is described in [2].

The C++ parallel implementation in `HiOp` is lightweight and portable since it is expressed and implemented only in terms of parallel (multi-)vector operations (implemented internally using BLAS level 1 and level 2 operations and MPI for communication) and BLAS level 3 and LAPACK operations for small dense matrices.

By using multithreadead BLAS and LAPACK libraries, *e.g.*, INTEL MKL, GotoBlas, Atlas, etc, additional, intra-node parallelism can be achieved. These libraries are usually machine/hardware specific and available for a variety of computer architectures. A list of BLAS/LAPACK implementations can be found at `https://en.wikipedia.org/wiki/Basic_Linear_Algebra_Subprograms#Implementations`.

2   Installing/building `HiOp`

`HiOp` is available on Lawrence Livermore National Laboratory (LLNL) github's page at `https://github.com/LLNL/hiop`. `HiOp` can be obtained by cloning the repository or by downloading the release archive(s). To clone from the repository, one needs to simply run

```
> git clone https://github.com/LLNL/hiop.git
```

2.1   Prerequisites

`HiOp` is written in C++ (C++98) and requires a C++ compiler. Also it requires BLAS and LAPACK, and, optionally, MPI. Not having MPI enabled in `HiOp` results in `HiOp` running on one processor only; still, significant multi-core parallelization can be obtained by using multithreaded BLAS and LAPACK. The CMake-based build system of `HiOp` generally detects these prerequisites automatically.

At this point the build system only supports Linux operating systems; it may also work on macOS. On the other hand, other than the build system, `HiOp` is platform independent and should run fine on Windows, macOS, and Linux.

2.2   Building, testing, and installing `HiOp`

The build system is based on CMake. Up-to-date detailed information about `HiOp` custom builds and installs are kept at `https://github.com/LLNL/hiop`.

A quick way to build and code is run the following commands in the 'build/' directory in the root `HiOp` directory:

```
> cmake ..
> make all
> make test
> make install
```

This will compile, build the static library and example executables, perform a couple of tests to detect potential issues during the installation, and will install `HiOp`'s header and the static library in the root directory under '_build_defaultDist/'

3   Interfacing with `HiOp`

Once `HiOp` is built, it can be used as the optimization solver in your application through the `HiOp`'s C++ interface and by linking with the static library. In this section we describe the steps necessary for doing this.

3.1 The mathematical formulation of the optimization problems solved by `HiOp`

In the spirit of NLP solvers, `Hiop` solves general nonlinear nonconvex optimization problems of the form

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{1}$$

$$\text{s.t.} \quad c(x) = c_E \qquad [y_c] \tag{2}$$

$$[v_l] \qquad d_l \leq d(x) \leq d_u \qquad [v_u] \tag{3}$$

$$[z_u] \qquad x_l \leq x \leq x_u \qquad [z_u] \tag{4}$$

Here $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^{m_E}$, $d : \mathbb{R}^n \to \mathbb{R}^{m_I}$. The bounds appearing in the inequality constraints (3) are assumed to be $d^l \in \mathbb{R}^{m_I} \cup \{-\infty\}$, $d^u \in \mathbb{R}^{m_I} \cup \{+\infty\}$, $d_i^l < d_i^u$, and at least of one of d_i^l and d_i^u are finite for all $i \in \{1, \ldots, m_I\}$. The bounds in (4) are such that $x^l \in \mathbb{R}^n \cup \{-\infty\}$, $x^u \in \mathbb{R}^n \cup \{+\infty\}$, and $x_i^l < x_i^u$, $i \in \{1, \ldots, n\}$. The quantities insides brackets are the Lagrange multipliers of the constraints. Whenever a bound is infinite, the corresponding multiplier is by convention zero.

The following quantities are required by `HiOp`:

D1 objective and constraint functions $f(x)$, $c(x)$, $d(x)$;

D2 the first-order derivatives of the above: $\nabla f(x)$, $Jc(x)$, $Jd(x)$;

D3 the simple bounds x_l and x_u, the inequalities bounds: d_l and d_u, and the right-hand size of the equality constraints c_E.

3.2 The C++ interface

The above optimization problem (1)-(4) can be specified by using the C++ interface, namely by deriving and providing an implementation for the `hiop::hiopInterfaceDenseConstraints` abstract class.

We present next the members of this abstract class that needs to be implemented in order to specify the parts D1-D3 of the optimization problem.

⚠ **Note:** All the functions that return `bool` should return `false` when an error occurs, otherwise should return `true`.

3.2.1 Specifying the optimization problem

All the methods of this section are "pure" virtual in `hiop::hiopInterfaceDenseConstraints` abstract class and need to be provided by the user implementation.

```
1 bool get_prob_sizes(long long& n, long long& m);
```

Provides the number of decision variables and the number of constraints ($m = m_E + m_I$).

```
1 bool get_vars_info(const long long& n, double *xlow, double* xupp,
2                    NonlinearityType* type);
```

Provides the lower and upper bounds x_l and x_u on the decision variables. When a variable (let us say the ith) has no lower or/and upper bounds, the ith entry of xlow and/or xupp should be less than $-1e20$ or/and larger than $1e20$, respectively. The last argument is not used and can set to any value of the enum `hiop::hiopInterfaceDenseConstraints::NonlinearityType`.

```
1 bool get_cons_info(const long long& m, double* clow, double* cupp,
2                    NonlinearityType* type);
```

Similar to the above, but for the inequality bounds d_l and d_u. For equalities, set the corresponding entries in clow and cupp equal to the desired value (from c_E).

```
1 bool eval_f(const long long& n,
2            const double* x, bool new_x,
3            double& obj_value);
```

Implement this method to compute the function value $f(x)$ in `obj_value` for the provided decision variables x. The input argument `new_x` specifies whether the variables x have been changed since the previous call of one of the `eval_` methods. Use this argument to "buffer" the objective and gradients function and derivative evaluations when this is possible.

```
1 bool eval_grad_f(const long long& n,
2                 const double* x, bool new_x,
3                 double* gradf);
```

Same as above but for $\nabla f(x)$.

```
1  bool eval_cons(const long long& n, const long long& m,
2                const long long& num_cons,
3                const long long* idx_cons, const double* x,
4                bool new_x, double* cons);
```

Implement this method to provide the value of the constraints $c(x)$ and/or $d(x)$. The input parameter `num_cons` specifies how many constraints (out of m) needs to evaluated; `idx_cons` array specifies the indexes, which are zero-based, of the constraints and is of size `num_cons`. These values should be provided in `cons`, which is also an array of size `num_cons`.

```
1 bool
2 eval_Jac_cons(const long long& n, const long long& m,
3               const long long& num_cons, const long long* idx_cons,
4               const double* x, bool new_x,
5               double** Jac);
```

Implement this method to provide the Jacobian of a subset of the constraints $c(x)$ and/or $d(x)$ in `Jac`; as above this subset is specified by the array `idx_cons`. To set the (i, j) entry of the Jacobian to v, one can use `Jac[i][j]=v;`. Alternatively, if you have a contiguous row-wise array storage of the Jacobian, simply (mem)copy it in `Jac[0]`.

3.2.2 Specifying a starting point for the optimization process

The user can provide an initial (primal) point x_0 for `HiOp`'s numerical process by implementing the method `get_starting_point` of the specification interface `hiop::hiopInterfaceDenseConstraints`.

```
1 bool get_starting_point(const long long& n, double* x0);
```

This method should return `true` if the initial point was successfully set. If the method returns `false`, then `HiOp` will use an all-zero starting point. The user does not have to implement this method, a default implementation that returns `false` is provided in the base class.

⚠ **Note:** The starting point returned in `x0` is subject to internal adjustments in `HiOp` and may differ from `x` with which the methods of the previous section are first called.

3.2.3 Specifying the inter-process/memory distribution of the problem

`HiOp` uses *data parallelism*, meaning that the data [D1]-[D3] of the optimization problem is distributed across processes (MPI ranks). It is **crucial** to understand the data distribution scheme in order to use `HiOp`'s interface properly.

   The general rule of thumb is to distribute any data of the problem with storage depending on n, namely the decision variables x and their bounds x_l and x_u, the gradient $\nabla f(x)$, and the Jacobians $Jc(x)$ and $Jd(x)$. The Jacobians, which are assumed to be dense matrices with n columns, are distributed column-wise.
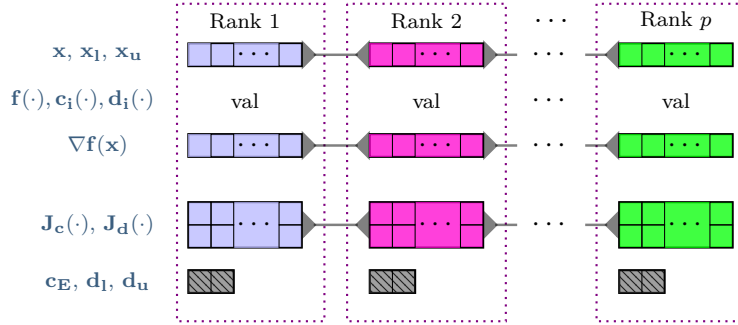

Figure 1: Depiction of the distribution of the data of the optimization problem (1)-(4) across MPI ranks. The vectors and matrices with storage dependent on the number of optimization variables are distributed. Other data, *i.e.*, scalar function values or vectors of small size (shown in dashed dark grey boxes), are replicated on each rank.

⚠ **Note:** All the `eval_` functions of the C++ interface provides local array slices of the above mentioned distributed data to the application code that implements `HiOp`'s C++ interface. The size of these local slices is the "local size" (specified by the application code through the `get_vecdistrib_info` method explained below) and is different from the "global size" n and parameter `n` of methods.

⚠ **Note:** Since the Jacobians are distributed column-wise, the implementer should populate the `Jac` argument of `eval_Jac_cons` with the "local" columns.

   On the other hand, the problem's data that does not have storage depending on n, is not distributed; instead, it is replicated on all ranks. Such data consist of c_E, d_l, d_u and the evaluations of $c(x)$ and $d(x)$.

```
1 bool get_MPI_comm(MPI_Comm& comm_out) ;
```

Use this method to specify the MPI communicator to be used by `HiOp`. It has a default implementation that will provide `MPI_COMM_WORLD`.

```
1 bool get_vecdistrib_info(long long global_n, long long* cols);
```

Use this method to specify the data distribution of the data of the problem that has storage depending on n. `HiOp` will call the implementation of this methods to obtain the partitioning/distribution of an hypothetical vector of size `global_n` across the MPI ranks. The array `cols` is of dimension number of ranks plus one and should be populated such that `cols[r]` and `cols[r+1]-1` specify the start and end indexes of the slice stored on rank r in the hypothetical vector. It has a default implementation that will returns `false`, indicating that `HiOp` should run in serial.

⚠ **Note:** `HiOp` also uses `get_vecdistrib_info` to obtain the information about the Jacobians' distribution across MPI ranks (this is possible since they are column-wise distributed).

Examples of how to use these functions can be found in the standalone drivers in `src/Drivers/` under the `HiOp`'s root directory.

3.2.4 Calling `HiOp`

Once an implementation of the `hiop::hiopInterfaceDenseConstraints` abstract interface class containing the user's NLP representation is available, the user code needs to create a `HiOp` problem formulation that encapsulate the NLP representation, instantiate an optimization algorithm class, and start the numerical optimization process. Assuming that the NLP representation is implemented in a class named `NlpEx1` (deriving `hiop::hiopInterfaceDenseConstraints`), the aforementioned sequence of steps can be performed by:

```
1  #include "NlpEx1.hpp"                   //the NLP representation class
2  #include "hiopNlpFormulation.hpp"       //HiOP encapsulation of the NLP
3  #include "hiopAlgFilterIPM.hpp"         //solver class
4  using namespace hiop;
5  ...
6  NlpEx1 nlp_interface();                              //instantiate your NLP representation↩
         class
7  hiopNlpDenseConstraints nlp(nlp_interface); //create HiOP encapsulation
8  nlp.options.SetNumericValue("mu0", 1e−2);  //set initial value for  barrier ↩
         parameter
9  hiopAlgFilterIPM solver(&nlp);                      //create a solver object
10 hiopSolveStatus status = solver.run();       //numerical optimization
11 double obj_value = solver.getObjective();   //get objective
12 ...
```

Various output quantities of the numerical optimization phase (*e.g.*, the optimal objective value and (primal) solution, status of the numerical optimization process, and solve statistics) can be retrieved from `HiOp`'s `hiopAlgFilterIPM` solver object. Most commonly used such methods are:

```
1 double getObjective() const;
2 void getSolution(double* x) const;
3 hiopSolveStatus getSolveStatus() const;
4 int getNumIterations() const;
```

The standalone drivers in `src/Drivers/` under the `HiOp`'s root directory contain more detailed examples of the use of `HiOp`.

3.3 Compiling and linking your project with the `HiOp` library

`HiOp`'s build system offers `HiOp` as a static library. Therefore, for a straightforward integration of `HiOp` in the user's project, one needs to

- append to the compiler's include path the location of the HiOP's headers:

  `-Ihiop-dir/include`

- specify `libhiop.a` to the linker, possibly adding the HiOP's library directory to the linker's libraries paths:

  `-Lhiop-dir/lib -lhiop`

Here, `hiop-dir` is the `HiOp`'s distribution directory (created using `HiOp`'s build system, in particular by using `make install` command).

4 Solver options

`HiOp` v0.1 prints all the available user options and their values on the standard output.

⚠ Each option i. should be of one of types numeric/double, integer, and string; ii. has a value associated; iii. possess a range of values; and, iv. has a default value.

The user can set `HiOp`'s options in two ways:

- via `hiop.options` file that should be placed in the same directory where the application driver using `HiOp` is executed. The format of the `hiop.options` is very basic, each of its lines should contain a single pair `option_name option_value`. Lines that begin with '#' are discarded. The option value is checked to have the correct type (numeric, integer, or string) and to be in the expected range. If the checks fail, then the option is set to the default value and a warning message is displayed.

- at runtime via the `HiOp`'s API using the `options` member of the `hiop::hiopInterfaceDenseConstraints` class. This object has three methods that allows the user to set options based on their types:

```
1   bool SetNumericValue(const char* name, const double& value);
2   bool SetIntegerValue(const char* name, const int& value);
3   bool SetStringValue (const char* name,  const char* value);
```

⚠ **Note:** Options set at runtime overwrite options set in `hiop.options` file.

4.1 List of available options

Name	Type	Default	Range	Explanation
tolerance	numeric	1e-08	1e-14 – 0.1	Absolute error tolerance for the NLP
max_iter	integer	3000	1 to 1000000	Max number of iterations

acceptable_tolerance	numeric	1e-06	1e-14 − 0.1	HiOp will terminate if the NLP residuals are below this option for `acceptable_iterations` many consecutive iterations
acceptable_iterations	integer	15	1 − 1000000	Number of iterations of acceptable tolerance after which HiOp terminates
verbosity_level	integer	3	0 − 12	Verbosity level: 0 no output (only errors), 1=0+warnings, 2=1 (reserved), 3=2+optimization output, 4=3+scalars; larger values explained in `hiopLogger.hpp`
mu0	numeric	1.0	1e-06 − 1000	Initial log-barrier parameter mu
kappa_eps	numeric	10.0	1e-06 − 1000	mu is reduced when when log-bar error is below kappa_eps × mu
kappa_mu	numeric	0.2	1e-08 − 0.999	linear reduction coefficient for mu (eqn. (7) in [5])
theta_mu	numeric	1.5	1.0 − 2.0	exponential reduction coefficient for mu (eqn. (7) in [5])
tau_min	numeric	0.99	0.9 − 0.99999	fraction-to-the-boundary parameter used in the line-search to back-off from the boundary (eqn. (8) in [5])
kappa1	numeric	1e-02	1e-08 − 1	sufficiently-away-from-the-boundary projection parameter used in the shift of the user-provided initial point
kappa2	numeric	1e-02	1e-08 − 0.49999	shift projection parameter used in initialization for double-bounded variables
dualsInitialization string	string	"lsq"	"lsq" "zero"	type of update of the multipliers of the eq. cons.
dualsUpdateType string	string	"lsq"	"lsq" "linear"	type of update of the multipliers of the eq. cons.
secant_memory_len	integer	6	0 − 256	length of the memory of the Hessian secant approximation

5   Licensing and copyright

HiOp is free software; you can modify it and/or redistribute it under the terms of the following modified BSD 3-clause license:

6   Acknowledgments

References

[1] R. H. Byrd, J. Nocedal, and R. B. Schnabel. Representations of quasi-newton matrices and their use in limited memory methods. *Mathematical Programming*, 63(1):129–156, 1994.

[2] C. G. Petra. A memory-distributed quasi-Newton solver for nonlinear programming problems with a small number of general constraints. Technical Report LLNL-JRNL-739001, Lawrence Livermore National Laboratory, October 2017.

[3] A. Wächter and L. T. Biegler. Line search filter methods for nonlinear programming: Local convergence. *SIAM Journal on Optimization*, 16(1):32–48, 2005.

[4] A. Wächter and L. T. Biegler. Line search filter methods for nonlinear programming: Motivation and global convergence. *SIAM Journal on Optimization*, 16(1):1–31, 2005.

[5] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.