Dimple User Manual
MATLAB Interface

Version 0.07

October 24, 2014

Contents

1 What is Dimple?

Dimple is an open-source API for probabilistic modeling and inference. Dimple allows the user to specify probabilistic models in the form of graphical models (factor graphs, Bayesian networks, or Markov networks), and performs inference on the model using a variety of supported algorithms.

Probabilistic graphical models unify a great number of models from machine learning, statistical text processing, vision, bioinformatics, and many other fields concerned with the analysis and understanding of noisy, incomplete, or inconsistent data. Graphical models reduce the complexity inherent in complex statistical models by dividing them into a series of logically (and statistically) independent components. By factoring the problem into sub-problems with known and simple interdependencies, and by adopting a common language to describe each sub-problem, one can considerably simplify the task of creating complex probabilistic models. A brief tutorial on graphical models can be found in Appendix A.

An important attribute of Dimple is that it allows the user to construct probabilistic models in a form that is largely independent of the algorithm used to perform inference on the model. This modular architecture benefits those who create probabilistic models by freeing them from the complexities of the inference algorithms, and it benefits those who develop new inference algorithms by allowing these algorithms to be implemented independently from any particular model or application.

Key features of Dimple:

- Supports both undirected and directed graphs.
- Supports a variety of *solvers* for performing inference, including sum-product and Gaussian belief propagation (BP), min-sum BP, particle BP, discrete junction tree, linear programming (LP), and Gibbs sampling.
- Supports both discrete and continuous variables.
- Supports arbitrary factor functions as well as a growing library of standard distributions and mathematical functions.
- Supports nested graphs.
- Supports rolled-up graphs (repeated HMM-like structures).
- Growing support for parameter estimation (including the EM algorithm).
- Supports both MATLAB[1] and Java[2] API.

Using MATLAB to program Dimple requires an existing licensed version of MATLAB, of version at least 2013b.

[1] MATLAB is a registered trademark of The Mathworks, Inc.

[2] Java is a registered trademark of Oracle and/or its affiliates.

2 Installing Dimple

2.1 Installing Binaries

Users can follow these instructions to install Dimple.

1. MATLAB of at least version 2013b is required.
2. Download the latest version of Dimple from http://dimple.probprog.org.
3. Extract the Dimple zip file.
4. Execute the startup.m script in the resulting Dimple directory to load Dimple in MATLAB.
5. To avoid having to manually change to this directory and execute this script every time you start MATLAB, you will need to add the following lines to MATLAB's startup.m file:

   ```
   cd <Path-to-Dimple>
   startup
   ```

   Google "MATLAB startup.m" for more details regarding startup.m files.
6. Verify the installation:
   (a) Start MATLAB
   (b) At the MATLAB command prompt type:

   ```
   testDimple;
   ```

   (c) Verify the output ends with something like the following (showing that all tests passed):

   ```


   **********************************************************************
   PASSED ALL TESTS
   252 of 252 tests passed, 0 failed
   **********************************************************************

   --testDimple
   ======================================================================
   ```

2.2 Installing from Source

Developers interested in investigating Dimple source code, helping with bug fixes, or contributing to the source code can install Dimple from source. Developers only interested in using Dimple should install from binaries (described in the previous section).

1. Download the source from https://github.com/AnalogDevicesLyricLabs/dimple
2. Install Gradle from http://www.gradle.org/. (Gradle is a Java build tool that pulls down jars from Maven repositories.)
3. Change to root directory
4. Run gradle by typing "gradle"

If you want to edit java files with Eclipse:

1. From eclipse, Import->Existing Projects Into Workspace
2. Browse to the dimple directory, select sovers/java, and click Finish.

2.3 Adjusting MATLAB's Java Memory Limit

Each object in Dimple corresponds to underlying Java objects. The amount of heap memory reserved for Java (when called from MATLAB) is limited, and typically low. In some cases, this can cause Dimple to fail if the memory it requires exceeds this modest limit. To increase the value of this limit, edit the file java.opts in the MATLAB startup directory, and add the following two lines:

```
-Xmx1024m
-Xms512m
```

The value after Xmx is the maximum amount of heap memory allocated to Java, and Xms is the starting value. You may use wish to use larger values if your system has sufficient memory. Google "MATLAB java.opts file" to determine the specific location of this file on your operating system.

3 Getting Started: Basic Examples

The following sections demonstrate Dimple with four basic examples. The first example is a simple hidden Markov model. The second models a 4-bit XOR constraint. The third demonstrates how to use nested graphs. The final example is a simple LDPC code.

See /demo/12_Tutorial for the code.

3.1 A Hidden Markov Model

We consider a very simple hidden Markov model (HMM), the Rainy/Sunny HMM illustrated in the Wikipedia article about HMMs. Two friends who live far apart, Alice and Bob, have a daily phone conversation during which Bob mentions what he has been doing during the day. Alice knows that Bob's activity on a given day depends solely on the weather on that day, and knows some general trends about the evolution of weather in Bob's area.

Alice believes she can model the weather in Bob's area as a Markov chain with two states 'Sunny' and 'Rainy'. She remembers hearing that on the first day of last week it was quite likely (70% chance) that it was sunny in Bob's town. She also knows that a sunny day follows another sunny day with 80% chance, while a sunny day succeeds a rainy day with 50% chance.

She knows Bob pretty well, and knows that Bob only ever does one of three things: stay inside to read a book, go for a walk, or cook. She knows that if it is sunny, Bob will go for a walk with 70% probability, cook with 20% probability, and stay inside with 10% probability. Conversely, if it is rainy, Bob will go for a walk with 20% probability, cook with 40% probability, and stay inside with 40% probability.

Bob told Alice that he went for a walk on Monday, Tuesday, and Thursday, cooked on Wednesday and Friday, and read a book on Saturday and Sunday.

Alice wants to know what the most likely weather is for every day of last week. The above naturally defines an HMM which can easily be modeled within Dimple

Creating the Factor Graph

The first thing to do in many Dimple programs, is to create a factor graph. This is easily done by instantiating a FactorGraph.

See demo/12_Tutorial/DimpleTutorial_HMM.m

```
HMM = FactorGraph();
```

Declaring Variables

Once a Factor Graph has been defined, we can define the variables of the factor graph. In our case, there will be seven variables, MondayWeather to SundayWeather. The syntax to create a variable is Discrete(domain,dimensions). In our case, the domain will consist of the two distinct values: Sunny and Rainy.For now, we will not specify dimensions (this will be covered in 4 Bit XOR example). The domain should either be a cell, or a matrix of numbers.

```
Domain={'sunny','rainy'};
MondayWeather=Discrete(Domain);
TuesdayWeather=Discrete(Domain);
WednesdayWeather=Discrete(Domain);
ThursdayWeather=Discrete(Domain);
FridayWeather=Discrete(Domain);
SaturdayWeather=Discrete(Domain);
SundayWeather=Discrete(Domain);
```

IMPORTANT: In the above, had we used the declaration Domain=['sunny','rainy'] instead (square brackets instead of curly brackets), the domain would have consisted of 10 letters instead of 2 strings (i.e., the variables would have been 10-ary instead of binary).

Adding Factors

We now add the different factors of the factor graph. We will first add the factors corresponding to the Markov Chain structure. This is done with addFactor, which is a method of the factor graph previously defined.

The method has syntax addFactor(funchandle,arguments), where funchandle is the handle of a regular MATLAB function (which can be specified by the user), and the arguments are the variables involved in the factor being defined (in the same order as the inputs of the MATLAB function). The number of inputs of the function referred to by the function handle has to be equal to the number of arguments of addFactor.

In our case, we define a simple MC_transition_Tutorial(state1,state2) as follows (See /demos/12_Tutorial/MC_transition_Tutorial.m):

```
function [probability]=MC_transition_Tutorial(state1,state2)

switch state1
    case 'sunny'
        if state2=='sunny'
            probability=0.8;
        else
            probability=0.2;

        end

    case 'rainy'
        probability =0.5;
end
```

We can now add the factor to the factor graphs by using the addFactor method:

```
HMM.addFactor(@MC_transition_Tutorial,MondayWeather,TuesdayWeather);
HMM.addFactor(@MC_transition_Tutorial,TuesdayWeather,WednesdayWeather);
HMM.addFactor(@MC_transition_Tutorial,WednesdayWeather,ThursdayWeather);
HMM.addFactor(@MC_transition_Tutorial,ThursdayWeather,FridayWeather);
HMM.addFactor(@MC_transition_Tutorial,FridayWeather,SaturdayWeather);
HMM.addFactor(@MC_transition_Tutorial,SaturdayWeather,SundayWeather);
```

We now need to add factors corresponding to the observations of each day. As it happens, when using an addFactor method, the arguments need not be all random variables—some can be declared as constants. We see now how to use this fact to easily add the observations to each day. We first declare an observation function

see /demo/12_Tutorial/observation_function_Tutorial.m:

```
function [probability]=observation_function_Tutorial(state,observation)

switch state
    case 'sunny'

        switch observation

            case 'walk'
                probability=0.7;
            case 'book'
                probability=0.1;
            case 'cook'
                probability=0.2;
        end

    case 'rainy'

        switch observation

            case 'walk'
                probability=0.2;
            case 'book'
                probability=0.4;
            case 'cook'
                probability=0.4;
        end
end
```

Adding the observations is then very easy:

```
HMM.addFactor(@observation_function_Tutorial,MondayWeather,'walk');
HMM.addFactor(@observation_function_Tutorial,TuesdayWeather,'walk');
HMM.addFactor(@observation_function_Tutorial,WednesdayWeather,'cook');
HMM.addFactor(@observation_function_Tutorial,ThursdayWeather,'walk');
HMM.addFactor(@observation_function_Tutorial,FridayWeather,'cook');
HMM.addFactor(@observation_function_Tutorial,SaturdayWeather,'book');
HMM.addFactor(@observation_function_Tutorial,SundayWeather,'book');
```

As we can see, though the function itself depends on two variables, each factor only depends on one random variable (the other argument being set as a constant during the addFactor call). This in effect creates a factor connected only to one variable of the factor graph.

There is a cost incurred every time addFactor is called in MATLAB. For large Factor Graphs with many factors of the same type, it can be advantageous to use the MATLAB vectorized version of addFactor. The following code will create a transition factor between each pair of consecutive variables and will execute much more quickly than a for loop with addFactor.

```
N = 1000;
Weather = Discrete(Domain,N,1);
HMM.addFactorVectorized(@MC_transition_Tutorial,Weather(1:(end-1)),Weather
    (2:end));
```

This works with continuous variables and Nested Graphs as well. In addition, it's possible to specify which dimensions to vectorize over:

```
N = 20;
b = Bit(3,N);
fg = FactorGraph();
fg.addFactorVectorized(@xorDelta,{b,2});
```

The previous code will create N xor factors across each of the N sets of 3 variables.

Specifying Inputs

The last step consists in adding the prior for the MondayWeather variable. We could, as above, use a factor with a single variable. Let us introduce a new property to easily add a single variable factor— the input property (on variables).

For a vector of probabilities (i.e., nonnegative numbers which sums up to one), Variable.Input adds a single factor which depends on this variable only, with values equal to those given by the vector.

In our case, we do:

```
MondayWeather.Input=[0.7 0.3];
```

The Input property can be used in several different ways. Some notes of interest regarding the Input property:

- The Input method can typically be used for prior probabilities of the root variables in a Bayes Net, or for the initial node of a Markov Chain or HMM.
- The Input property can also be used for any factors with only one variable, for instance, for observation factors (see the introduction to factor graphs on how to remove observations and turn them into single node factors).

- The observation_function_Tutorial in the above example was not entirely required (see below)—we could have used the input property instead.
- IMPORTANT: Unlike the addFactor method, the Input property can be used only once for each variable. That is, once you have specified an input for a variable, re-specifying this input will destroy the previous factor and create a new one. In the example above, using only the Input property, it would not have been possible to separately incorporate both the prior on Monday's weather and the factor corresponding to Mondays observation. However, this feature is very useful when Input is used to specify external information, and when the user wants to see the effect of external information. Say for instance that Bob mentions to Alice that it rained on Wednesday. Alice can simply use the call WednesdayWeather.Input=[0 1] . If Bob corrects himself and says he misremembered, and that it actually was sunny that day, Alice can correct the information using again the call WednesdayWeather.Input=[1 0] .

Solving the Factor Graph

Finally, we explain how to solve the factor graph by using the solve, iterate, and NumIterations factor graph methods.

The typical way of solving a factor graph will be by choosing the number of iterations and then running the solver for the desired number of iterations. In our case, this is simply done by typing the following code:

```
HMM.NumIterations=20;
HMM.solve;
```

IMPORTANT:

By default, the solver will use either a Flooding Schedule or a Tree Schedule depending on whether the factor-graph contains cycles. A loopy graph (one containing at least one cycle) will use a Flooding Schedule by default. This schedule can be described as:

- Compute all variable nodes
- Compute all factor nodes
- Repeat for a specified number of iterations

If the factor-graph is a tree (one that contains no cycles), the solver will automatically detect this and use a Tree Schedule by default. In this schedule, each node is updated in an order that will result in the correct beliefs being computed after just one iteration. In this case, NumIterations should be 1, which is its default value.

Accessing Results

Once the solver has finished running, we can access the marginal distribution of each variable by using the Belief property:

```
belief  = TuesdayWeather.Belief;
```

This returns a vector of probability of the same length as the domain total size (i.e., the product of its dimensions), with the probability of each domain variable. Another way to solve the factor graph is to use the Solver.iterate(n) method, which runs the factor graph for n iterations (without arguments, it runs for 1 iteration).

```
HMM.Solver.iterate();
HMM.Solver.iterate(5);
```

IMPORTANT: The iterate method is useful to access intermediate results (i.e, see how beliefs change through the iterations).

IMPORTANT: One distinction between the solve and iterate methods is that all messages and beliefs are reinitialized when starting the solve method. Running solve twice in a row is therefore identical to running it once, unlike iterate. When calling iterate() for the first time, first call initialize(), which reinitialize all messages.

3.2 A 4-bit XOR

The following example creates a simple factor graph with 4 variables tied through a 4-bit XOR, with ‘priors’ (we abuse language here and call ‘prior’ the probability distribution of each random variable if they were not tied through the 4-bit XOR).

Through this example, we will learn how to define arrays of random variables, see how to use MATLAB indexing within Dimple, see an example of a hard constraint in a factor graph, and see how to use the Bit type of random variable.

We consider a set of four random variables (B1,B2,B3,B4) taking values 0 or 1. The joint probability distribution is given by:

$$Pr(B_1, B_2, B_3, B_4) = \delta_{B_1+B_2+B_3+B_4=0(mod2)}P(B_1)P(B_2)P(B_3)P(B_4)$$

where the delta function is equal to 1 if the underlying constraint is satisfied, and 0 otherwise (this ensures that illegal assignment have probability zero). The $P(B_i)$ are single variable factors which help model which values of B_i are more likely to be taken (we call them ‘priors’, though, once again, this is an abuse of language. Typically, the factor will represent an observation of O_i of variable B_i, and the factor $P(B_i)$ is set equal to the normalized function $P(O_i|B_i)$ [3]

For our example, we will choose $P(B_1 = 1) = P(B_2 = 1) = P(B_3 = 1) = .8$ and $P(B_4 = 1)$ = 0.5.

We will build our factor graph in two different ways. The first directly uses a 4-bit XOR, and uses mostly tools seen in the previous example, while the second introduces the Bit kind of random variable, and how to use an intermediate variable to reduce the degree of a factor graph with parity checks (i.e., XOR function).

The first way of building the factor graph uses an inefficient N-bit XOR function defined as follows

From /demo/12_Tutorial/BadNBitXorDelta_Tutorial.m:

```
function [valid]=BadNBitXorDelta_Tutorial(variables)

valid=mod(sum(variables),2)==0;

end
```

Using everything we have learned in the previous example, the sequence of instructions we use is simply:

From /demo/12_Tutorial/DimpleTutorial_BadNBitXor.m:

[3] Normalizing $P(O_i|B_i)$ happens to be equal to $P(B_i|O_i)$ in a factor graph with only the two variables O_i and B_i with a prior on both values of B_i being equally likely.

```
FourBitXor=FactorGraph();
Domain=[0;1];
B1=Discrete(Domain);
B2=Discrete(Domain);
B3=Discrete(Domain);
B4=Discrete(Domain);
FourBitXor.addFactor(@BadNBitXorDelta_Tutorial, [B1,B2,B3,B4]);
B1.Input=[0.2 0.8];
B2.Input=[0.2 0.8];
B3.Input=[0.2 0.8];
B4.Input=[0.5 0.5];
FourBitXor.solve;
disp(B1.Value);
disp(B1.Belief);
```

Note that the MATLAB BadNBitXorDelta_Tutorial is a function which takes only ONE argument, but this argument is an array. This is reflected in the declaration FourBitXor.addFactor (@BadNBitXorDelta_Tutorial, [B1,B2,B3,B4]), where we created an array of random variables using square brackets.

IMPORTANT: We also introduce the Discrete method ‘Value’, which returns the most likely assignment of that random variable.

The first remark above is rather important, as it highlights of the concept of arrays of random variables in the MATLAB Dimple API. Dimple handles arrays of random variables in a very natural manner, and most array indexing operations of MATLAB are supported in Dimple in a similar fashion.

To create a multidimensional array of random variables in the MATLAB Dimple API, we construct a Variable class instance where every constructor argument after the first is a dimension of an array:

```
VarArray=Variable(Domain,dimension1,dimension2,.., dimensionk)
```

creates an array of variables with domain ‘Domain’, and with dimensions [dimension1][dimension2]..[dimensionk].

IMPORTANT: If only one dimension is specified, Dimple creates a square array.

```
VarArray1=Discrete(Domain,3,1)
VarArray2=Discrete(Domain,3)
VarArray3=Discrete(Domain,3,3)
```

In the example above, VarArray1 is a vector of 3 random variables with domain Domain, while VarArray2 and VarArray2 are both a 3-by-3 matrix of random variables with domain Domain.

Another way to create an array is, as above, to group the variables using square brackets:

```
RowArray=[ B1 B2 B3 B4];
ColumnArray=RowArray';
```

IMPORTANT: The above method is a simple ”grouping” (reference) of variables ; it does not duplicate them.

Concatenation and subindexing of random arrays work in exactly the same fashion as MATLAB.

```
SubArray1=VarArray2(2,1:2);
SubArray2=VarArray2(1,2:3);
SubArray3=[SubArray1;SubArray2];
```

repmat works as well. The following code snippet will set C to a 10x10 Bit matrix. Rather than creating new variables, it simply replicates the existing variables.

```
b = Bit(10,1);
c = repmat(b,1,10);
```

Using the Variable methods Belief, Value or Domain on a random array returns the array of Beliefs (resp. ML values, domains, etc..) for these variables.

```
SubArray2.Value
```

returns a (1,2) array containing the most likely values of VarArray2(1,2) and VarArray2(1,3).

IMPORTANT: Since Beliefs or Inputs are arrays themselves, calling them on an array returns an array one dimension larger.

Often, we will find it useful to have random Bits. For that purpose, one can directly create random Bits with the Bit class. A Bit is simply a Discrete with Domain [0, 1]. Also, in order to simplify Input declarations, for a Bit variable named BitVar, BitVar.Input takes a single number, the probability of 1.

Similarly, BitVar.Belief returns the marginal probability of BitVar being equal to 1.

The second version of the 4-bit XOR will uses Bit variables to represent our random variables. Furthermore, it will decompose the 4 bits XOR into two 3-bits XOR. It is indeed easy to prove that $B_1 + B_2 + B_3 + B_4 = 0(mod2)$ is equivalent to

$$\begin{aligned} B_1 + B_2 + C &= 0(mod2) \\ B_3 + B_4 + C &= 0(mod2) \end{aligned}$$

for an internal bit C. While the reduction from one 4-bit to two 3-bit XORs might not seem tremendous, it is easy to see that more generally, any N-bit XOR can be reduced to a tree of 3-bit XORs, with depth log(N). Because the complexity of running Belief Propagation is exponential in the degree of factors, using this technique leads to dramatic speed improvements.

Using all the techniques mentioned above, we derive a new factor graph for the 4-bit XOR, equivalent to the one previously defined.

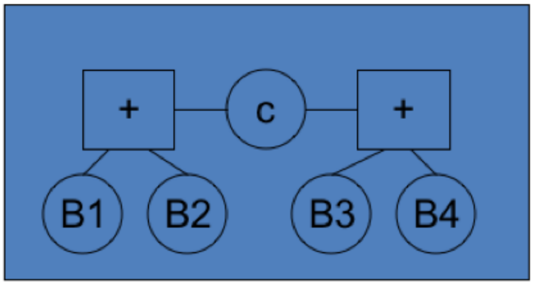
From /demo/12_Tutorial/xorDeltaTutorial.m:

```
function valid = XorDeltaTutorial(x,y,z)
    valid = bitXor(bitXor(x,y),z) == 0;
end
```

From demo/12_Tutorial/DimpleTutorial_4BitXor.m:

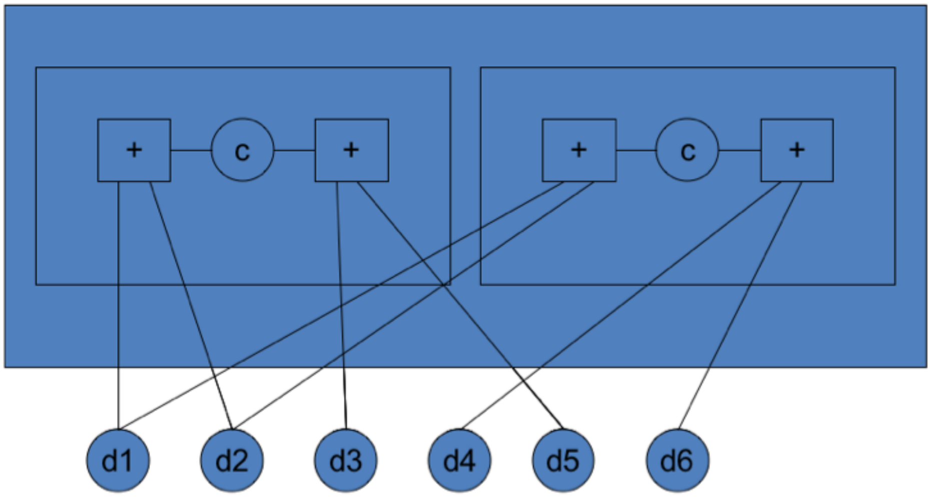
```
XorGraph = FactorGraph();
b = Bit(4,1);
c = Bit();
XorGraph.addFactor(@XorDeltaTutorial,b(1),b(2),c);
XorGraph.addFactor(@XorDeltaTutorial,b(3),b(4),c);
b.Input = [ .8 .8 .8 .5];
XorGraph.solve();
disp(b.Belief);
disp(b.Value);
```

The following figure represents the Factor Graph that is created and solved in this example.



3.3 Nested Graphs

Suppose we wish to use two 4-bit XORs from the previous example to create a 6-bit code. The following diagram shows our desired Factor Graph.



Dimple provides a way to replicate multiple copies of a Factor Graph and nest these instances in a containing Factor Graph. A nested Factor Graph can be seen as a special factor function between a set of variables ('connector variables'), which, when 'zoomed in,' is in fact another factor graph, with factors involving both the connector variables, and other 'internal variables.' In the second version of the XOR example, we created a 4-bit XOR connected to 4 variables, by also creating an extra Bit C. What if we wanted to use that construction as a 4-bit XOR for potentially many different sets of 4 bits, overlapping or not, by replicating the factor graph as needed? Nested factor graphs provide an elegant solution to this problem.

A nestable factor graph is created by specifying first a set of "connector" random variables, and instantiating a FactorGraph with these variables passed in to the constructor.

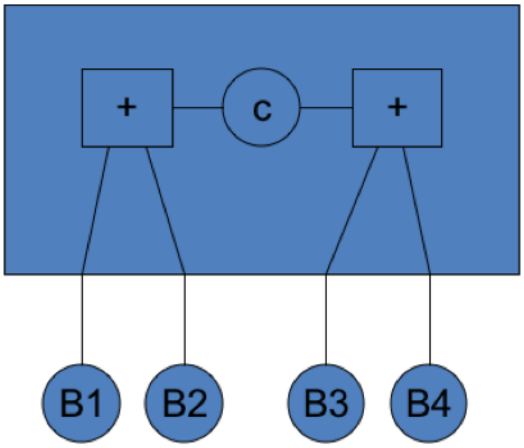
IMPORTANT: The factor graph defined this way is still a proper factor graph, and in principle, we can run BP on it. However, in practice, it is used as a "helper" factor graph (and are "helper" random variables), which will mostly be replicated in the actual factor graph of interest.

The following code creates a nestable factor graph, with connector variables (B_1, B_2, B_3, B_4)

From /demo/12_Tutorial/DimpleTutorialNested.m:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Define 4 bit xor from two 3 bit xors
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
b = Bit(4,1);
XorGraph = FactorGraph(b);
c = Bit();
XorGraph.addFactor(@xorDeltaTutorial,b(1),b(2),c);
XorGraph.addFactor(@xorDeltaTutorial,b(3),b(4),c);
```

IMPORTANT: In principle, variables attached to a Factor Graph can be defined before or after defining the factor graph (but obviously always before the factors they are connected to). However, connector variables naturally need to be defined before the nestable factor graph.



Consider a nestable factor graph NestableFactorGraph(connectvar1, connectvar2,..,vark) with k connector variables. Consider also an actual factor graph of interest, FactorGraph, containing (among others) k variables of interest (var1,..,vark). By using the addFactor method, we can replicate the NestableFactorGraph and use it to connect the variables (var1,..,vark) in the same way the connector variables are connected in the nestable factor graph: FactorGraph.addFactor(NestableFactorGraph,var1,..,vark).

In essence, the nestable factor graph represents a factor between the dummy connector variables, and the addFactor method is adding this factor to the desired actual variables.

IMPORTANT: Nested factor graphs support arbitrary levels of nesting. That is, a nested factor graph can be composed of nested factor graphs.

Armed with this tool, we can very simply use our custom 4-bit XOR to implement the factor graph described at the beginning of the section:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Create graph for 6 bit code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
d = Bit(6,1);
MyGraph = FactorGraph(d);
MyGraph.addFactor(XorGraph,d([1:3 5]));
MyGraph.addFactor(XorGraph,d([1 2 4 6]));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Set input and Solve
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
d.Input = [.75 .6 .9 .1 .2 .9];
MyGraph.NumIterations = 20;
MyGraph.solve();
disp(d.Value');
```

IMPORTANT: Note the use of standard MATLAB array indexing in the d([1:3 5]) method above. Also, note that technically, our nestable factor graph declared above only had one argument (not four), an array of four random variables (as opposed to four separately declared random variables). As a result, the addFactor call requires a single argument—an array of four random

3.4 An LDPC Code

For our last example, we will show how to use a library function to very quickly code up a complex factor graph.

We present the code that defines a Factor Graph to decode an LDPC (low-density parity check) code (in the file /demo/11_SingleCodewordLDPC/createLDPC.m):

```
function [ldpc,x] = createLdpc()
    A = load('matrixout.txt');
    blockLength = size(A,2);
    numCheckEquations = size(A,1);
    ldpc = FactorGraph();
        x = Bit(blockLength,1);
        for i = 1:numCheckEquations
            varIndices = find(A(i,:));
            gd = getNBitXorDef(length(varIndices));
            ldpc.addFactor(gd,x(varIndices));
        end
     ldpc.NumIterations = 100;
end
```

At this point, all the code above should be clear, except for the getNBitXorDef call. getNBitXorDef(N) returns a nestable factor graph with an array of N bits as connected variables, representing an N-bit XOR factor between those N bits. It implements the most efficient (in some sense) implementation of an N-bit XOR using a tree of 3-bits XORs.

The check matrix connectivity is loaded from a file into matrix A. The for loop iterates all check equations, creates a nestable n-bit XOR Graph and adds that graph to the LDPC. The user can set Input on x and the returned variables, solve the ldpc object, and retrieve the results with Value or Beliefs.

4 How to Use Dimple

4.1 Defining Models

4.1.1 Overview of Graph Structures

While Dimple supports a variety of graphical model forms, including factor graphs, Bayesian networks, or Markov networks, the Dimple programming interface is oriented toward the language of factor graphs. Factor graphs are normally defined to be bipartite graphs—that is, graphs consisting of two types of nodes: *factor nodes* and *variable nodes*, connected by *edges*. Factor nodes connect only to variable nodes and vice versa. A mathematical overview of factor graphs can be found in Appendix A.

The model creation aspect of Dimple primarily involves defining variables and connecting these variables to factors. While dimple supports basic factor graphs, in which variables and factors are connected directly with no hierarchy, Dimple also supports more complex structures such as *nested* factor graphs, and *rolled-up* factor graphs, described below.

Basic factor graph: This is a graph in which all variables and factors are connected directly, with no hierarchy. Basic graphs are inherently finite in extent.

Nested factor graph: When a portion of a factor graph is used more than once, it may be convenient to describe that portion once, and then place copies of that sub-graph within a larger graph. Nested graphs can be used as a form of modularity that allows abstracting the details of a sub-graph from the description of the outer graph. A sub-graph can be thought of as a single factor in the outer graph that happens to be described as a factor graph itself. In Dimple, graphs may be nested to an arbitrary depth of hierarchy. Creation of a nested sub-graph within another graph is described in section 4.1.4.5.

Rolled-up factor graph: In some cases, a structure in a factor graph is repeated a great many times, or an unbounded or data-dependent number of times. In such cases, Dimple allows creation of rolled-up factor graphs, where only one copy of the repeated section is described in the model. The result is a factor graph that is implicitly unrolled when inference is performed. The amount of unrolling depends on a potentially unbounded set of data sources. Creation of rolled-up factor graphs is described in section 4.3.

4.1.2 Creating a Graph

To create a new factor graph, call the FactorGraph constructor and assign it to a variable. For example,

```
myGraph = FactorGraph();
```

Note that Dimple makes use of MATLAB objects. In this case, a FactorGraph is such an object, and “FactorGraph()” corresponds to the constructor of that object; in this case, called with no arguments (if there are no arguments, the parentheses may be omitted). The FactorGraph object has a number of properties and methods that can be called using MATLAB’s standard notation for object properties and methods. For example, “myGraph.Name” refers to the Name property of the factor graph.

The created factor graph does not yet contain any variables or factors. Once a graph is created, then variables and factors can be added to it, as described in the subsequent sections.

Creating a factor graph in this way corresponds to a basic factor graph, or the top-level graph in a nested-graph hierarchy. To create a factor graph that will ultimately be used as a sub-graph in a nested graph hierarchy, the constructor must include arguments that refer to the “boundary variables” that will connect it to the outer graph. For example,

```
mySubGraph = FactorGraph(varA, varB, varC);
```

In this case, the variables listed in the constructor, must already have been created. Creation of variables is described in section 4.1.3.

Creation of sub-graphs is described in more detail in section 4.1.4.5.

4.1.3 Creating Variables

4.1.3.1 Types of Variables

Dimple supports several types of variables:

Discrete: A Discrete variable is one that has a finite set of possible states. The domain of a discrete variable is a list of the possible states. These states may be anything—numbers, strings, arrays, objects, etc.

Bit: A Bit is a special-case of a discrete variable that has exactly two states: 0 and 1. Note that when using Bit variables, there are some differences in the API versus using Discrete variables. These differences are noted in the appropriate sections of this manual.

Real: A Real variable is a continuous variable defined on the real line, or some subset of the real line.

RealJoint: A RealJoint variable is a multidimensional continuous variable, defined on $\mathbb{R}^{\mathbb{N}}$ or a subset of $\mathbb{R}^{\mathbb{N}}$. Unlike a vector of Real variables, the components of a RealJoint variable cannot be connected independently in a graph.

Complex: A Complex variable is a special-case of a RealJoint variable with two dimensions. In some cases, the values of a Complex variable are expressed as MATLAB complex numbers.

FiniteFieldVariable: A FiniteFieldVariable is a special-case of a discrete variable that represents a finite field with $N = 2^n$ elements. These fields are often used, for example, in error correcting codes. These variables can be used along with certain custom factors that are implemented more efficiently for sum-product belief propagation than the alternative using discrete variables and factors implemented directly. See section 4.4 for more information on how these variables are used.

A factor graph may mix any combination of these variable types, though there are some limitations in the use of certain variable types by some solvers. Specifically, some solvers do not currently support Real variables.

A variable of a given type can be created using the corresponding object constructor. For example,

```
myDiscreteVariable = Discrete(1:10);
myBitVariable = Bit();
myRealVariable = Real();
```

In this case, each of the above lines created a single variable of the corresponding type. In the case of a Discrete variable, a constructor argument is required, which defines the domain of the variable. In the above example, the domain is the set of integer numbers 1 through 10.

Note that variables may be created prior to creation of the factor graph that they will ultimately become part of[4]. A variable is not yet associated with any factor graph until it is connected to at least one factor, as described in section 4.1.4.

4.1.3.2 Specifying Variable Domains

4.1.3.2.1 Discrete Variables

When creating a Discrete variable, the domain of that variable must be specified[5]. Once created, the domain of a variable cannot change.

The domain of the discrete variable is always the first argument of the constructor. The domain is a list in the form of either a MATLAB numerical array, or a MATLAB cell array. Each entry in the array corresponds to one possible state of the variable. In the case of a

[4] For nested graphs, at least the boundary variables *must* be created prior to creation of the factor graph.

[5] Note that for Bit variables, the domain is implicit and is not specified in the constructor.

cell array, the entries are arbitrary–they may be, for example, strings, vectors, arrays, or objects.

Some examples:

```
v1 = Discrete(1:10);
v2 = Discrete([1.5, 1/3, 27.4327, -13.6, sqrt(2), sin(pi/7)]);
v3 = Discrete({’Sun’, ’Clouds’, ’Rain’, ’Snow’});
v4 = Discrete({ [1 2; 3 4], [5 6 7; 8 9 10], 3.7});
```

Even though the domains are discrete, the actual values of the domain states may be anything, including real numbers. Though, when using real numbers and objects, care must be taken, for example, when making equality comparisons.

The values of a domain need not be defined in-line in the constructor, but may be defined elsewhere in the program. For example:

```
myDomain = 1:10;
...
myVariable = Discrete(myDomain);
```

Alternatively, the domain can be defined as a DiscreteDomain object, which provides an object wrapper around the domain. For example:

```
myDomain = DiscreteDomain(1:10);
...
myVariable = Discrete(myDomain);
```

In this case, the myDomain object has properties that can be queried, such as “.Elements”, which provides a list of the elements of the domain.

In this case, the myDomain object has properties that can be queried, such as “.Elements” , which provides a list of the elements of the domain.

To see the domain of a variable, you can query the Domain property. For a discrete variable, this returns a DiscreteDomain object, which you can query as described above to see the elements of the domain. For example:

```
disp(myVariable.Domain.Elements);
```

The Domain property could be used, for example, to take one variable and create others that have the same domain. For example,

```
newVariable = Discrete(otherVariable.Domain);
```

4.1.3.2.2 Real Variables

When creating a Real variable, specifying a domain is optional. If no domain is specified, the domain is assumed to be the entire real line from $-\infty$ to ∞.

The domain of a Real variable is an array of two real numbers: the first specifies the lower bound on the variable, and the second specifies the upper bound (the upper bound must be greater than or equal to the lower). Either or both of the values may be $-\infty$ or ∞.

As with discrete variables, the domain may be specified ahead of time, and may be created by defining in this case a RealDomain object, which takes two arguments: the lower and upper bound, respectively. Some examples:

```
r1 = Real();
r2 = Real([ 0 Inf]);
r3 = Real([-pi pi]);
myDomain = RealDomain(-2.6, 12.4);
r4 = Real(myDomain);
```

The Domain property of a Real variable returns in this case a RealDomain object, which contains two properties, LB and UB, which correspond to the lower bound and upper bound, respectively.

4.1.3.3 Creating Arrays of Variables

In many cases, a factor graph contains a set of variables that all have the same type and the same domain. In this case, it is often convenient to treat these as a single vector or multidimensional array of variables. An array of variables can be created all at once by specifying the dimensions of the array as the final arguments in the constructor. For example,

```
v1 = Bit(10,10);
v2 = Discrete(myDomain, 3, 4, 5, 2, 10);
v3 = Real([-pi pi], 20, 1);
```

The first example creates a 10x10 matrix, the second a multidimensional array with 5 dimensions, each with the specified length, and the third a column vector.

The dimensions follow the MATLAB convention of row, followed by column, followed by any additional dimensions. Also like other MATLAB functions, specifying a single dimension of length N implies a square NxN matrix.

Once a variable array has been created, individual variables or sub-arrays may be referenced using standard MATLAB notation for array indexing. For example:

```
oneVariable = v1(3,2);
subArray = v2(:,:,:,1,8:end);
subVector = v3(5:10);
```

4.1.3.4 Naming Variables

Variables can be named, which can be useful in debugging as well as displaying a factor graph visually. A single variable can be named by setting the Name property:

```
myVariable.Name = 'My variable';
```

The variable name can be retrieved by referencing the Name property.

An array of variables can be named, each with distinct names by using the setNames method[6]:

```
myVariableArray.setNames('X');
```

This results in a unique name assigned to each variable with the specified prefix, followed by a unique vector variable index number.

All variables in a factor graph must have unique names. Sometimes it's desirable to give multiple variables the same label for plotting or displaying. In this case, users can set the Label property. If the Label property is not set, plotting and displaying uses the name. If the Label property is set, the Label is used when displaying and plotting.

```
myVar.Label = 'myvar';
```

4.1.4 Creating Factors and Connections to Variables

4.1.4.1 Basic Factor Creation

Creation of a factor graph involves defining the variables of that graph, and then successively creating factors that connect to these variables.

Before creating a factor, the following must already have been created:

[6] For large arrays, using this method is significantly faster than naming each variable individually in a loop.

- The factor-graph into which the factor will be placed.
- All of the variables that the factor will connect to.

The most basic way of creating a factor is by calling the "addFactor" method on the factor graph object. For example:

```
myGraph.addFactor(factorSpecification, myVariable1, myVariable2);
```

Resulting from this method call, the specified factor is created, the factor is connected to the variables listed in the arguments, and the factor and all connected variables become part of the designated factor graph (if they are not already).

The factor specification is one or more arguments that define the factor. Dimple supports a variety of ways of specifying a factor, each of which is described in more detail in subsequent sections. A factor specification may be one of the following:

- A MATLAB function handle
- A factor-table
- A built-in factor
- The name of a built-in factor
- A sub-graph
- A built-in overloaded MATLAB operator[7]
- A custom Java FactorFunction object[8]

After the factor specification argument(s), the remaining arguments are all treated as variables or constants. The variables may be individual variables or arrays of variables, or combinations thereof.

The number and type of variables that can be connected to a factor depend on the type of factor being created. In some cases, a factor is defined flexibly to accommodate an arbitrary number of connected variables, while in other cases there are restrictions. In most cases, the order of variables matters. The definition of a factor generally requires a specific order of variables, where each variable or set of variables may be required to be of a specific type. In such cases, the arguments of the addFactor method call must appear in the required order.

For example, "Normal" is a built-in factor that describes a set of normally distributed variables with variable mean and precision. In this case, the first argument must be the mean, the second must be the precision, and the remaining one or more variables are the normally distributed variables. In this case, the factor creation could look like the following:

[7] This method of factor creation does not use addFactor, but instead uses a different syntax to create the factor implicitly.

[8] The mechanism to create and use custom Java FactorFunction objects is described in Appendix B.

```
fg = FactorGraph();
meanValue = Real();
precisionValue = Real([0 Inf]);
normalSamples = Real(100,100);
fg.addFactor(’Normal’, meanValue, precisionValue, normalSamples);
```

In this example, the factor specification argument is the name of the built-in factor, and the normalSamples argument refers to the entire array of 100x100 normally distributed variables. The number of these sample variables is of arbitrary length only because of the way the “Normal” built-in factor was defined.

Note that these variables could have been listed explicitly as separate arguments. For example, if there were only two such variables, we could have written:

```
normalSamples = Real(2,1);
fg.addFactor(’Normal’, meanValue, precisionValue, normalSamples(1),
    normalSamples(2));
```

When creating a factor, it is sometimes convenient to supply a constant value to one or more of the factor’s arguments instead of a variable. This can be done simply by substituting a value that is not a Dimple variable. In this case, the value must be consistent with the possible values the particular factor is designed to accept. In the above example, if we wished to have a variable mean value, but define a fixed precision of 0.2, we could have written:

```
fg.addFactor(’Normal’, meanValue, 0.2, normalSamples(1), normalSamples(2));
```

Note that in this case, the “Normal” built-in factor requires the precision to be positive, so if we had provided a constant value of -0.2, for example, this would have resulted in an error. The particular requirements of a factor are specific to the definition of that factor.

4.1.4.2 Vectorized Factor Creation

In many cases, a factor graph has repeated patterns of variables and factors, where the factors are all of the same type. In this case, Dimple provides a vectorized factor creation method that allows adding an entire array of factors all at once with a single operation. Using this approach is typically significantly faster than adding factors one-by-one in a loop.

The most straightforward use of this method is when all of the associated variables are all arrays with the same dimensions as the array of factors. For example, supposed we have a 100x100 array of variables, A, and a 100x100 array of variables, B. And we wish to add a set of 100x100 factors, where each factor connects the corresponding A and B variables. In this case, we would write:

```
A = Bit(100,100);
B = Bit(100,100);
fg.addFactorVectorized(factorSpecification, A, B);
```

This need not be limited to arrays with two dimensions, but generalizes to an arbitrary set of dimensions.

The vectorized approach can also be used to connect arrays of variables to individual variables or constants. For example, continuing the above example:

```
C = Real();
fg.addFactorVectorized(factorSpecification, A, B, C);
fg.addFactorVectorized(factorSpecification, A, B, 0.1);
```

In the first addFactorVectorized call, above, each of the 100x100 factors created is connected to the corresponding element of A and B, and all of them connect to the variable C. In the second case, above, all of the 100x100 factors use the constant 0.1 as the third argument.

Note that unlike the previous use of addFactor, having array arguments resulted in many copies of the factors rather than a single factor connecting to many variables. In some cases, it is desirable to have some combination of these. Dimple provides a syntax for determining which dimensions of an array should be vectorized (creating multiple factors) and which should be treated as array inputs to a each factor.

Expanding on the example above, say A were a 100x100 array and B were a 100x100x5 array, and for each 100x100 factors we wished to connect one element of A with each of the 5 elements of B corresponding the entire length of its third dimension. In this case, we could write:

```
A = Bit(100,100);
B = Bit(100,100,5);
fg.addFactorVectorized(factorSpecification, A, {B, [1 2]});
```

In the above, the variable argument B is replaced with a cell-array containing the variable followed by an array listing the dimensions for which the "vectorized" connection to the array of factors should be made. In this case, first two dimensions, of size 100x100 correspond to the 100x100 factors that we wish to vectorize factor creation. Since dimension 3 was not included in this list, then for each value of row and column dimension, the entire length of B variables in the third dimension are connected to the corresponding factor.

All arguments passed to addFactorVectorized that are arrays of variables (rather than single variables or constants) must have the same number of dimensions over which they are vectorized and those dimensions must be of the same size. For example, in the above example, we could not have specified vectorizing over B's second and third dimensions, which do not have the same size as A. However, if B were instead 5x100x100, we could have done so.

If arrays of variables that must be connected are not already of the same size, or have the same order of dimensions, the MATLAB functions repmat, permute, reshape, or squeeze may be applied to variable arrays to reorganize them into the appropriate form. In the case of repmat, Dimple does not actually make multiple copies of the variables, but instead provides repeated references to the same variables. This allows, for example, a vector of variables to be connected to a grid of variables, where each element of the vector connects to an entire row of factors in the grid. Below is an example of such a case:

```
A = Bit(100,100);
B = Bit(100,1);
fg.addFactorVectorized(factorSpecification, A, repmat(B,1,100));
```

4.1.4.3 Using MATLAB Factor Functions

A factor may be specified by defining a MATLAB function, and passing a handle to that function. The function must accept values that correspond to the possible states of the connected variables (in the same order as specified in the addFactor call), and return a non-negative weight corresponding to the unnormalized value of the factor.

Using a MATLAB function to specify a factor is valid in Dimple only when all of the connected variables are discrete (either Discrete or Bit). Real values are allowed in other parts of a graph, as long as they are not directly connected to a factor defined this way.

Starting with a very simple example, suppose we wish to create a factor that corresponds to a two-input logical AND function, where the first argument is the output of the AND function, and the second and third are the inputs. This is a deterministic factor, where the weight of the factor is 1 (or any arbitrary constant) if the output variable equals the logical AND of the two inputs, and 0 otherwise.

To define this factor function, we create a MATLAB function that is accessible via the MATLAB path. In this example, we will call the function AndExample, in the file AndExample.m. We can define this function as follows:

```
function factorValue = AndExample(out, in1, in2)
    factorValue = (in1 && in2) == out;
end
```

We then create a factor using this factor function, connecting it to previously defined Bit variables, x, y, and z:

```
myFactorGraph.addFactor(@AndExample, x, y, z);
```

The “@” symbol indicates a MATLAB function handle, which is a reference to the previously defined function.

It would have been possible to instead define the function in-line as a MATLAB anonymous function. Though, this does not allow a function to be reused to create multiple factors. The following is equivalent to the example above:

```
myFactorGraph.addFactor(@(out, in1, in2) (in1 && in2) == out, x, y, z);
```

In this case, the factor function had a fixed number of arguments. If instead, we would like the number of input arguments to be variable, we can make use of MATLAB's variable argument lists:

```
function factorValue = AndExampleImproved(out, varargin)
    factorValue = prod(cell2mat(varargin)) == out;
end
```

We can use this factor with an arbitrary number of input variables:

```
myFactorGraph.addFactor(@AndExampleImproved, x, y, z, a, b, c, d);
```

MATLAB factor functions need not be so simple. First of all, their outputs can be any non-negative value. But they can also take more interesting values than just the integers. For example, suppose we define a factor of two variables over a domain consisting of strings such that the factor output value is proportional to the number of times the second string is found in the first, relative to the length of the first string.

```
function factorValue = StringMatchExample(string, pattern)
    factorValue = numel(strfind(string, pattern)) / length(string);
end
```

4.1.4.4 Using Factor Tables

When Dimple creates a factor using a factor function, for some solvers, behind the scenes it translates that factor function into a table by enumerating all possible states of all connected variables. While steps are taken to make this efficient, including storing only non-zero values and reusing tables for identical factors, the time it takes to create the factor table can in some cases be very large. In some situations, a user may have specific knowledge of the structure of a factor that would allow them to create the same table much more efficiently. To accommodate such cases, Dimple allows factors to be specified using user-created factor tables.

A factor table consists of two parts: a two dimensional array of integers and a single dimensional array of doubles. Each row of the two dimensional table represents a combination of variable values for which the factor value is non-zero. Each column represents a variable connected to the factor. The values of this table specify an index into the discrete domain of a variable. Each row of the two dimensional table corresponds to one entry of the array of doubles, where that entry contains the value of the factor corresponding to the corresponding set of variable values.

Once the user has created the table, they can create a factor using this table in one of two ways. The first is to provide the two dimensional array of indices and vector of values directly as the first two arguments of the addFactor call, respectively:

```
fg = FactorGraph();
b = Bit(2,1);
indices = [0 0; 1 1];
values = [2 1];
fg.addFactor(indices, values, b);
```

In the following example we first create a factor table object and then create a factor using that table. This has the advantage of using less overall memory if this same factor table will be used in multiple factors.

```
fg = FactorGraph();
b = Bit(2,1);
indices = [0 0; 1 1];
values = [1 1];
myFactorTable = FactorTable(indices, values, b.Domain, b.Domain);
fg.addFactor(myFactorTable, b);
```

4.1.4.5 Using Sub-Graphs

In a nested graph, a factor at one layer of the graph hierarchy can correspond to an entire sub-graph. To add a sub-graph as a factor to another graph, first the sub-graph must have already been created. A sub-graph is created almost the same as any ordinary graph, with the exception of defining a subset of its variables to be “boundary” variables. These indicate how the sub-graph will connect to other variables in the outer graph.

To understand how sub-graph creation works, we first note that when a sub-graph is added to an outer graph, a new copy of the sub-graph is made, with entirely new variables and factors. The original sub-graph is used only as a template for creating the copies. This means that the actual variables used in the sub-graph are never directly used in the final nested graph. Internal variables within the sub-graph are created new when the sub-graph is added. Boundary variables, on the other hand, are connected to variables in the outer graph, which might already exist in that graph.

When a sub-graph is created, its boundary variables must be defined in the graph constructor. The boundary variables listed in the constructor must be of the identical type and have the identical domain (in the case of discrete variables) as the variables they will later connect when added to the outer graph. Additionally, the order of variables listed in creation of the sub-graph must match exactly the order of variables listed when adding the sub-graph to an outer graph.

For example, we define a subgraph as follows:

```
a = Discrete(1:10);
b = Bit;
x = Bit;
mySubGraph = FactorGraph(a, b);
mySubGraph.addFactor(exampleFactor1, a, b);
mySubGraph.addFactor(exampleFactor2, b, x);
```

To add this subgraph to an outer graph, we use addFactor (or addFactorVectorized), specifying the factor using the subgraph object.

```
N = 5;
fg = FactorGraph;
P = Discrete(1:10, N, 1);
Q = Bit(N,1);
for i = 1:N
    fg.addFactor(mySubGraph, P(i), Q(i));
end
```

Equivalently, this can be vectorized using:

```
fg.addFactorVectorized(mySubGraph, P, Q);
```

4.1.4.6 Using Built-In Factors

Dimple supports a set of built-in factors that can be specified when adding a factor to a graph. The complete list of available built-in factors can be found in section 5.9.

In many cases, a built-in factor may be specified simply by referring to the name of the factor as a string (which is case sensitive). As an example,

```
Mean = Real();
Precision = Real();
Values = Real(1,100);
MyGraph.addFactor('Normal', Mean, Precision, Values);
```

Built-in factors may be specified by name only if no constructor arguments are needed. If constructor arguments are needed, then there are two ways to specify the factor. The preferred way is to create a FactorFunction object, which takes the name of the factor followed by the constructor arguments for that factor. For example:

```
MyGraph.addFactor(FactorFunction('Gamma', 1, 1), X);
```

The same FactorFunction can be used more than once, which avoids creating additional copies of the FactorFunction object. For example:

```
myFactorFunction = FactorFunction('Gamma', 1, 1);
MyGraph.addFactor(myFactorFunction, X1);
MyGraph.addFactor(myFactorFunction, X2);
```

A short-hand notation may alternatively be used, in which the name of the factor function and its constructor arguments are contained in a cell array. For example:

```
MyGraph.addFactor({'Gamma', 1, 1}, X);
```

4.1.4.7 Implicit Factor Creation Using Overloaded Operators and Functions

Dimple supports a set of built-in factors that can be added implicitly using overloaded MATLAB operators or functions. For example,

```
fg = FactorGraph();
a = Discrete(1:4);
b = Discrete(1:10);
c = a + b;
```

The last line of this example will automatically create a new variable, c, and a 'Sum' factor connected to variables c, a, and b. The domain of c will be defined appropriately given the domains of the input variables. In this example, the domain of c would automatically be set to the range [2:14].

These operations can be compounded in a single line of code, and variables of different data types as well as constants can be intermingled (as long as the type is supported by the specific operator). In this case, intermediate anonymous variables will be created in the graph associated with intermediate results of the operation. For example,

```
z = (a + b) * c^d - sqrt(-e);
```

Like using the addFactor method, adding factors implicitly can include constants. Specifically, for binary operators, one of the inputs may be a constant instead of a variable. For example:

```
x = a^2;
y = (a + b + 2) * 3;
```

Since adding a factor implicitly does not specifically refer to the factor graph, the graph to which these factors are added is also implicit. In particular, these implicitly defined factors are added to the last factor graph that was created. So, in the first example above, the factor would be added to fg, regardless of whether other factor graphs had previously been created.

Adding built-in factors implicitly using overloaded MATLAB operators or functions can also be vectorized, with some limitations. Specifically, if each of the input variables are vectors of the same dimension, then the result will be to create a vector of output variables of the same dimension, along with a vector of factors relating the inputs and outputs.

In some cases, to be consistent with MATLAB notation, there is a distinction made between the vectorized and non-vectorized operator. Specifically, Dimple uses MATLAB's notation for pointwise product and power operators to indicate a vectorized operation. For example, if variables a through e are vectors of variables of identical size, then the following would create a variable vector z, and a series of factors relating these variables.

```
z = (a .* b) + c.^d - sqrt(-e);
```

For binary operators, one of the inputs may be a scalar variable or a scalar constant instead of a variable vector. For a scalar variable, the result is that scalar variable connecting to each instance of the factors that are created. For a constant, each instance of the factor uses the same constant for that input (vectors of distinct constants are not currently supported). As an example:

```
a = Real();
b = Discrete(domain, 1, 10);
z = a + b;
```

In cases where the user wishes to use operator overloading with nested graphs, it may be necessary to use the FactorGraph's addBoundaryVariables method. For instance, if they create a nested graph such that $y = a + b$ and the user requires y to be a boundary variable, the addBoundaryVariables method must be used. For example:

```
ng = FactorGraph();
a = Bit();
b = Bit();
y = a+b;
ng.addBoundaryVariables(y,a,b);

fg = FactorGraph();
a2 = Bit();
b2 = Bit();
```

```
y2 = Discrete([0 1 2]);
fg.addFactor(ng,y2,a2,b2);
```

The specific set of operators supported is given in section 5.10.

4.1.4.8 Naming Factors

Just like variables, factors can be named, which can be useful in debugging as well as displaying a factor graph visually. To name a factor, the factor object must be accessible via a variable. When using addFactor, the result is the factor object, which can be assigned to a variable. A single factor can be named by setting the Name property of the factor object.

```
myFactor = fg.addFactor(exampleFactor, a, b, c);
myFactor.Name = ’My factor’;
```

And the factor name can be retrieved by referencing the Name property.

An array of factors can be named, each with distinct names by using the setNames method:

```
myFactorArray = addFactorVectorized(exampleFactor, a, b, c);
myFactorArray.setNames(’F’);
```

This results in a unique name assigned to each factor with the specified prefix, followed by a unique vector variable index number.

Just like Variables, Factors also support the notion of a Label, which can be used to allow multiple factors to share the same label.

4.1.5 Modifying an Existing Graph

4.1.5.1 Removing a Factor

It is possible to remove a Factor from an existing FactorGraph:

```
fg = FactorGraph();
   b = Bit(3,1);
   fg.addFactor(@xorDelta,b(1:2));
   f = fg.addFactor(@xorDelta,b(2:3));
   b.Input = [.8 .8 .6];
   fg.NumIterations = 2;
   fg.solve();
   assertElementsAlmostEqual([.96 .96 .96]’,b.Belief);
```

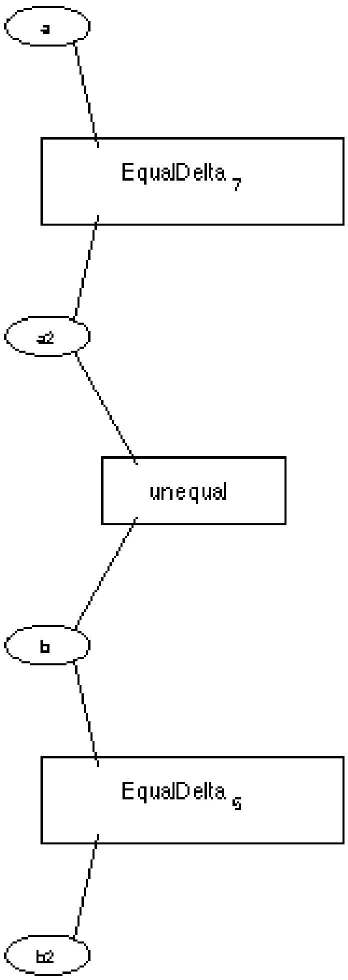
```
fg.removeFactor(f);
fg.solve();
p1 = .8*.8;
p0 = .2*.2;
total = p1+p0;
p1 = p1/total;
p0 = p0/total;
assertElementsAlmostEqual([p1 p1 .6]',b.Belief);
```

4.1.5.2 Splitting Variables

It can be useful to make a copy of a variable and relate it to the old variable with an equals factor. The following code shows how to do this.

```
a = Bit();
a.Name = 'a';
b = Bit();
b.Name = 'b';

fg = FactorGraph();

f = fg.addFactor(@(x,y) x~=y,a,b);
f.Name = 'unequal';

b2 = fg.split(b);
b2.Name = 'b2';
a2 = fg.split(a,f);
a2.Name = 'a2';

fg.plot(1);
```

We've added code to name all the variables and factors so that the following plot is informative.



Note that the split method takes a list of factors as the second through nth argument. This is the list of factors that will be moved from the original variable to the copied variable. All unspecified factors will remain connected to the initial variable.

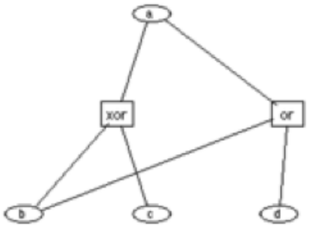
4.1.5.3 Joining Variables

It is possible to join variables in an existing graph, which will create a new joint variable and modify all factors connected to the original variables to reconnect to the new joint variable. This can be useful in eliminating loops in a graph. The following code creates a loopy graph and then uses join to remove the loop.

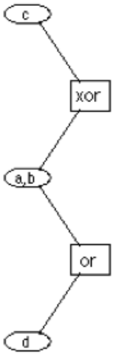
```
a = Bit();
a.Name = 'a';
```

```
b = Bit();
b.Name  = ’b’;
c = Bit();
c.Name = ’c’;
d = Bit();
d.Name = ’d’;

fg = FactorGraph();
f1 = fg.addFactor(@xorDelta,a,b,c);
f1.Name = ’xor’;
f2 = fg.addFactor(@(x,y,z) (x|y)==z ,a,b,d);
f2.Name = ’or’;

newvar = fg.join(a,b);
newvar.Name = ’a,b’;

fg.plot(1);
```

The following is the loopy graph:



And after joining the variables we have:

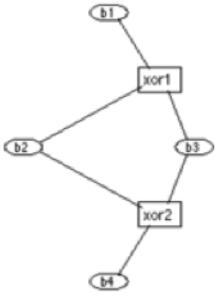


4.1.5.4 Joining Factors

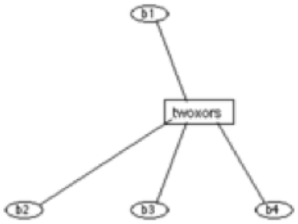
It is possible to remove loops by joining factors as well as by joining variables. (*NOTE: an easier way to eliminate loops is to use the Junction Tree solver (see 5.6.8), which will produce a transformed version of the graph without altering the original graph.*)

```
b = Bit(4,1);
for i = 1:4
    b(i).Name = sprintf('b%d',i);
end
fg = FactorGraph();
f1 = fg.addFactor(@xorDelta,b(1:3));
f2 = fg.addFactor(@xorDelta,b(2:4));

f3 = fg.join(f1,f2);
f3.Name = 'twoxors';

b.Input = input;
fg.solve();
actualBelief = b.Belief;

fg.plot(1);
```

The following plot shows the graph with the loops:



And the following plot shows the graph after the factor is joined:



To join factors, Dimple does the following:

- Find the variables in common between two factors.
- Take the cartesian product of the tables but discard rows where the common variable indices differ.
- Consolidate the columns with common variables.
- Multiply the values for each row.

4.1.5.5 Changing Factor Tables

For factors connected only to discrete variables, the factors are stored in the form of a factor table (regardless of how the factor was originally specified). It is possible to modify some or all of the entries of that factor table in an existing factor graph.

Modifying a single entry or set of entries in a factor table can be done by getting the FactorTable property and calling the set method. For example, the following changes a single entry in a factor. In this example, the factor has two edges, the first with domain containing strings, the second with a domain containing numbers. The arguments of the set method prior to the last one specify the domain settings for which the factor value is to be changed. The final argument is the new factor value for the particular domain setting, which will replace the previous value.

```
factor.FactorTable.set('red', 1, 0.4);
```

To change multiple elements at once, the sets of arguments may be included in a comma-separated list of cell arrays:

```
factor.FactorTable.set({'red', 1, 0.4}, {'blue', 1, 0.25}, {'blue', 2, 0});
```

More detail on the use of the set method can be found in section 5.3.4.3.1.

Alternatively, the entire factor table can be replaced with another one. For this, the change method would be used instead:

```
indexList = [0 0 0; 0 1 1; 1 1 0; 1 0 1];
weightList = [0.2 0.15 0.4 0.9];
factor.FactorTable.change(indexList, weightList);
```

Here the indexList is is an array where each row represents a set of zero-based indices into the list of domain elements for each successive domain in the set of domains associated with the variables connected to this factor. The weightList is a one-dimensional array of real-valued entries in the factor table, where each entry corresponds to the indices given by the corresponding row of the indexList.

IMPORTANT: Identical factor tables are automatically shared between factors. If changing the factor table for a factor and if that factor is shared by other factors, then this changes it globally for all such factors.

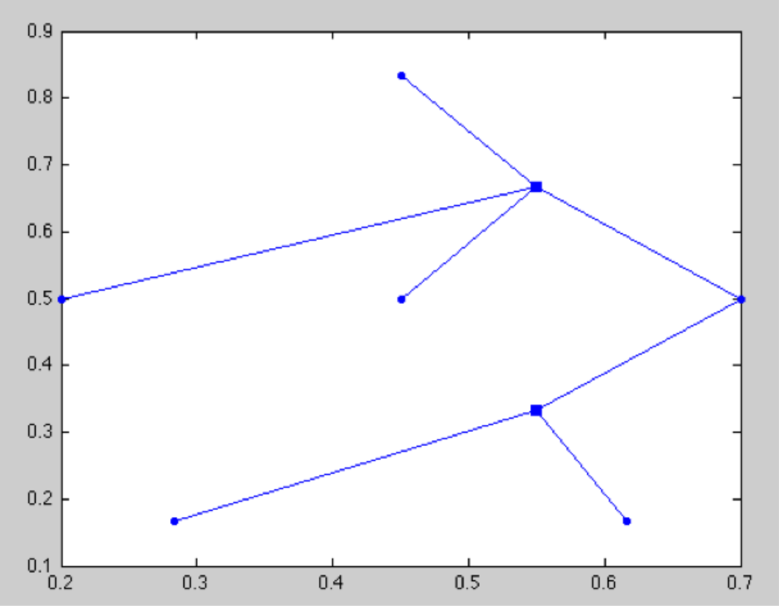
4.1.6 Plotting a Graph

When debugging Factor Graphs, it is sometimes useful to be able to plot a Factor Graph. The FactorGraph class provides a plot method that can be used to visualize a Factor Graph.

The following code describes how to use the plot function in various ways.

```
%First we build a Factor Graph to use for plotting examples
fg = FactorGraph();
b = Bit(6,1);
for i = 1:6
    b(i).Name = sprintf(’b%d’,i);
end

%We use Label rather than Name for the factors so that we can assign
%them the same Label.  Name must be a unique identifier,
%Label is just used for printing/plotting.
f1 = fg.addFactor(@xorDelta,b(1:4));
f1.Label = ’f’;
f2 = fg.addFactor(@xorDelta,b(4:6));
f2.Label = ’f’;


pause_time = 1;

%Calling plot with no arguments shows no labels.  It draws variables as
%circles and factors as squares.
fg.plot();
```

This will result in the following graph:



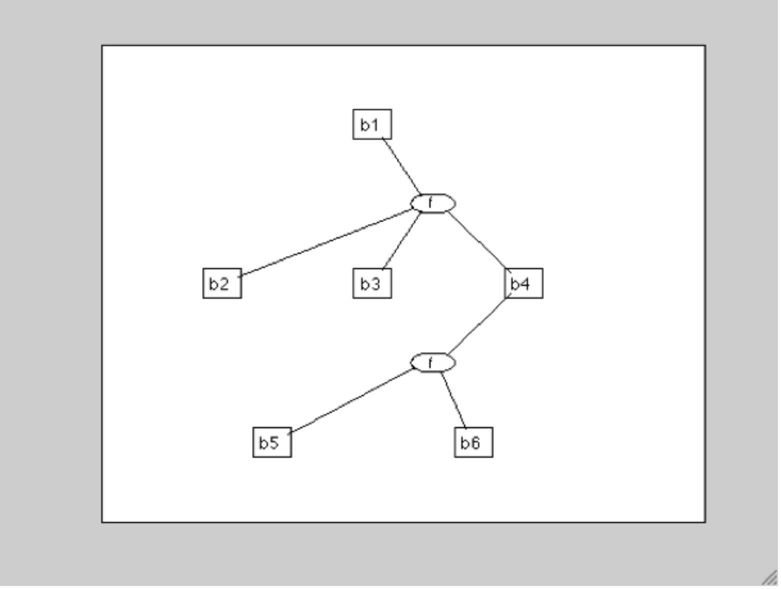
Note that factors are displayed as squares and variables as circles.

```
pause(pause_time);

%The following is equivalent to the previous plot command.  We are simply
%explicitly turning off labels.
fg.plot(’labels’,false);
```

Results in the same plot as above.

```
pause(pause_time);

%Now we turn on the labels.  Now we see the names we assigned to the
%various nodes and variables of the Factor Graph.
fg.plot('labels',true);
```

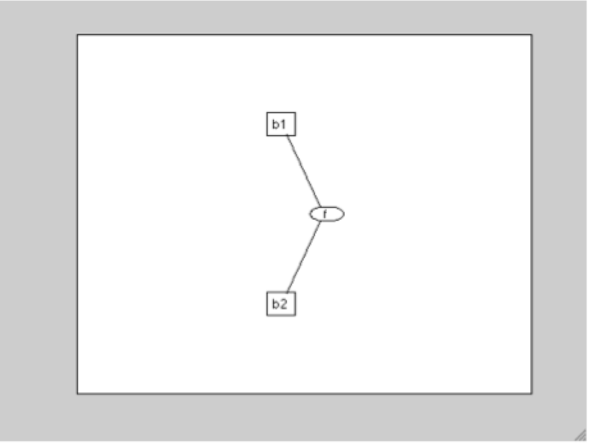
Results in the following:



If the user has specified a Label, those will be displayed, otherwise the object's Name's will be displayed.

```
pause(pause_time)

%We can specify a subset of nodes to plot
fg.plot('labels',1,'nodes',{b(1:2),f1});
```

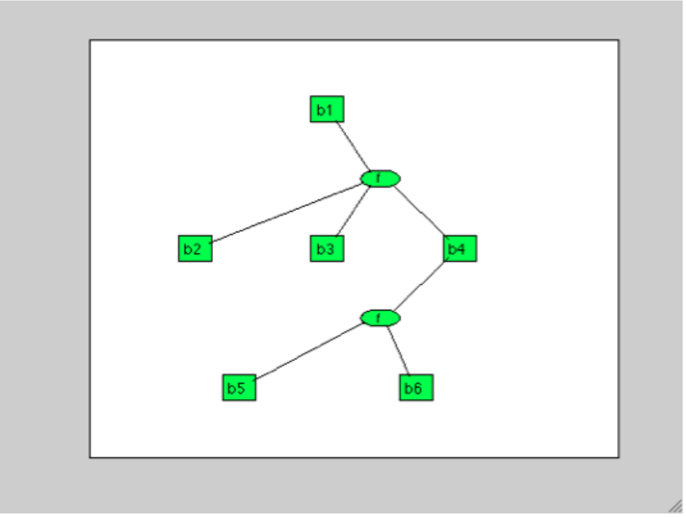
Results in the following:



Only the specified nodes and their connectivity were included.

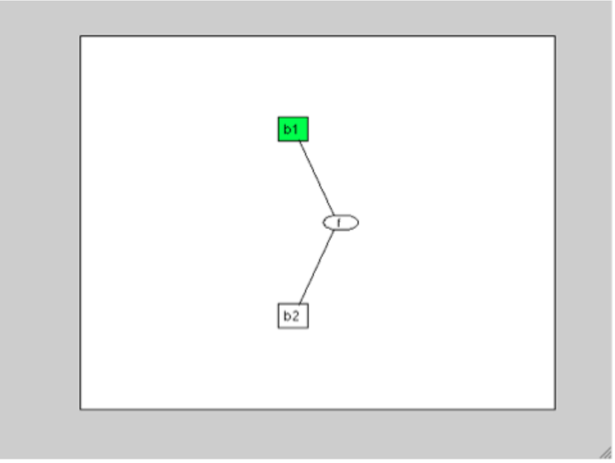
```
pause(pause_time)

%By can set a global color for all the nodes in the graph.
fg.plot('labels',1,'color','g');
```

Setting a global color:



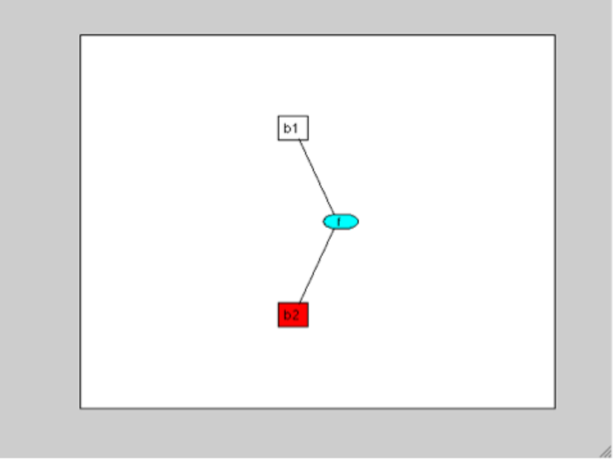
```
pause(pause_time)

%We can specify a color for one node in the graph.
fg.plot('labels',1,'nodes',{b(1:2),f1},'color',b(1),'g');
```

Setting a color for a specific node:



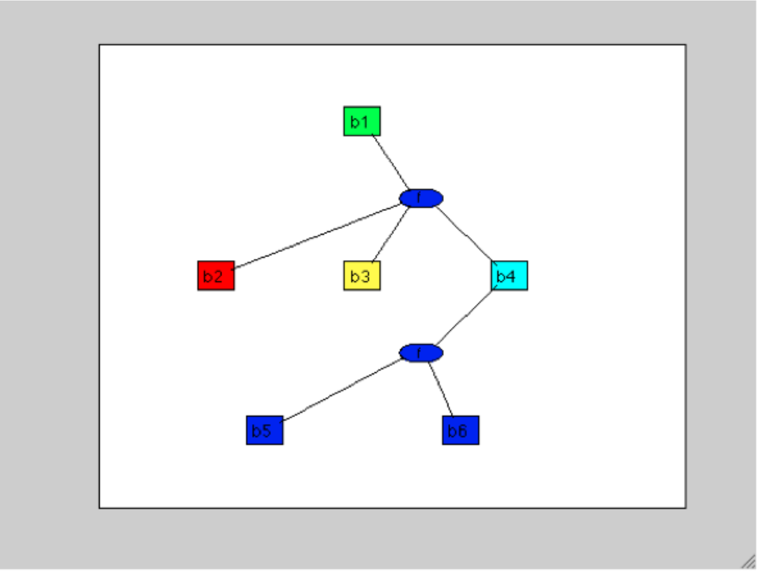
```
pause(pause_time)

%We can specify colors for multiple nodes in the graph.
fg.plot('labels',1,'nodes',{b(1:2),f1},'color',{b(2),f1},{'r','c'});
```

Setting colors for multiple nodes:



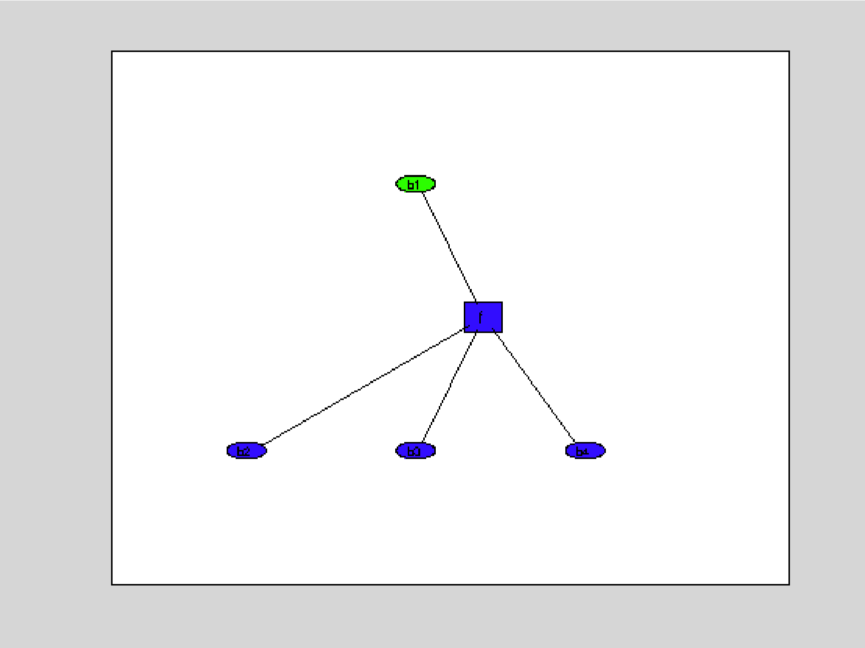
```
pause(pause_time)

%We can mix setting a global color, colors for a single node mutliple
%times, and colors for multiple nodes.  The global color is used in all
%cases where a color has not explicitly been set for a node.
fg.plot('labels',1,'color',b(1),'g','color',b(2),'r','color',{b(3),b(4)},{'y
    ','c'},'color','b');
```

Mixing and matching the various coloring options:



```
pause(pause_time)

%We can also specify a root node on which we perform a depth first search
%up to a specific depth and then only print nodes up to that depth.
for depth = 0:5

    %Furthermore, we color the root node green so we know which is the root
    %node.
    fg.plot('labels',1,'depth',b(1),depth,'color',b(1),'g','color','b');

    pause(pause_time);
end
```

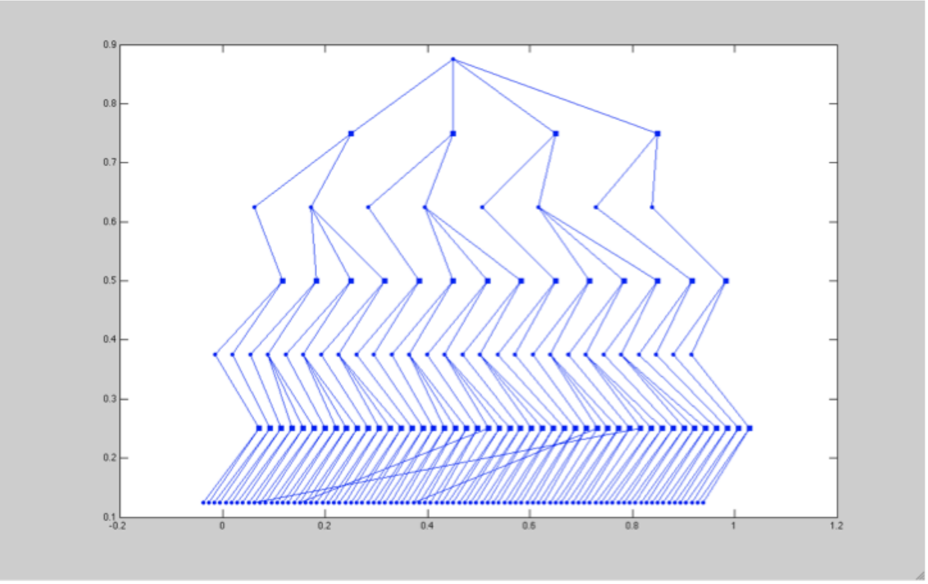
Specifying a depth will display a specified root node and all nodes that are N steps away. The following shows the result of calling plot with ‘depth’ of b(1) and 2:



We also colored the root node green to make it clear which was the root node.

```
%The following shows how using the depth feature we might be able to find
%out interesting information.  Here we increase the depth until we visually
%see a loop.
[ldpc,vars] = createLDPC();
v = vars(1);

for depth = 0:6
    ldpc.plot('depth',v,depth);

    pause(pause_time);
end
```

Here, we show how we can use plot to find interesting information about a large graph. The final plot with a depth of 6 shows a cycle in an LDPC Factor Graph:



4.1.6.1 Plotting Nested Graphs

By default, the plotting method ignores hierarchy and plots the flattened graph. If the user specifies the 'nesting' parameter, however, they can specify how deep to descend into the hierarchy before considering NestedGraphs to be Factors and plotting them as such.

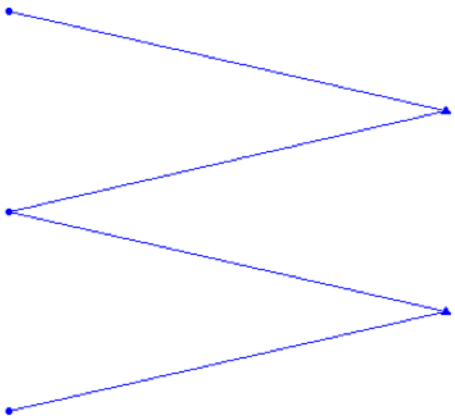
When labels are turned off, nested graphs are displayed as triangles

First let's build a graph with three levels of nesting.

```
b = Bit(2,1);
template1 = FactorGraph(b);
iv = Bit();
template1.addFactor(@xorDelta,b(1),iv);
template1.addFactor(@xorDelta,b(2),iv);

b = Bit(2,1);
template2 = FactorGraph(b);
iv = Bit();
template2.addFactor(template1,b(1),iv);
template2.addFactor(template1,b(2),iv);

template2.plot();

b = Bit(2,1);
fg = FactorGraph(b);
iv = Bit();
fg.addFactor(template2,b(1),iv);
fg.addFactor(template2,b(2),iv);
```

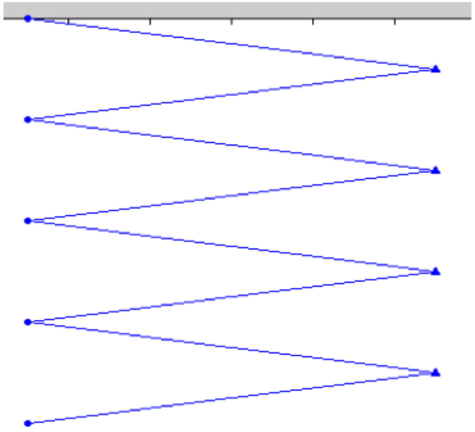
Here we show the graph plotted with various levels of nesting.

fg.plot('nesting',0);

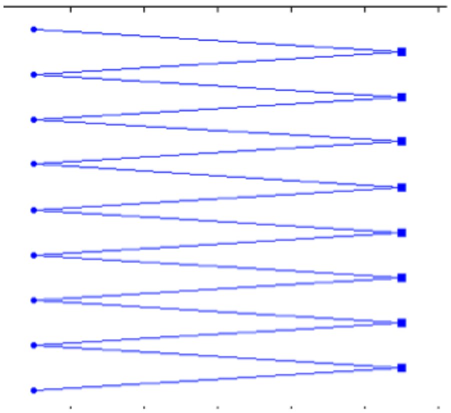


Notice the Nested Graphs show up as triangles.

```
fg.plot('nesting',1);
```



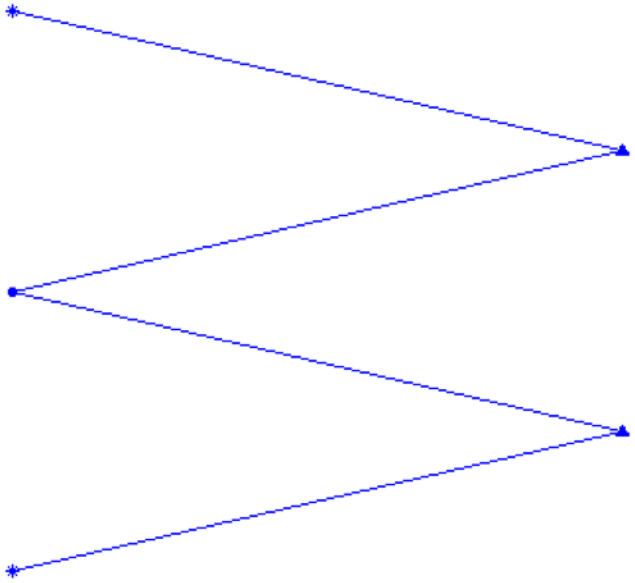
```
fg.plot('nesting',2);
```



Note that once we have reached the bottom, we are actually seeing the Factors plotted.

We can retrieve an instance of a nested graph and plot that with nesting set.

```
fg.NestedGraphs{1}.plot('nesting',0);
pause(pause_time);
```

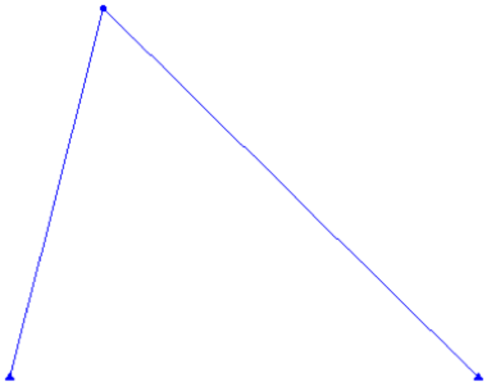


When plotting graphs, boundary variables show up as stars.

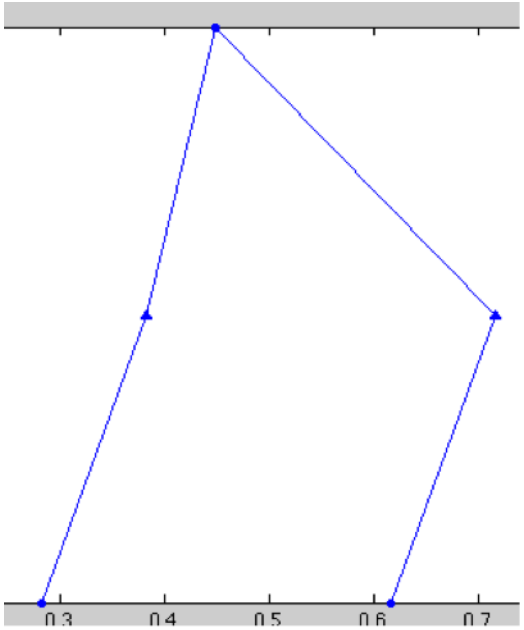
Next we mix depth first search with nesting

```
fg.plot('nesting',0,'depth',iv,0);
```

```
fg.plot('nesting',0,'depth',iv,1);
```



```
fg.plot('nesting',0,'depth',iv,2);
```



4.1.7 Structuring Software for Model Portability

Because the specification of a model in Dimple is expressed in code, this code can be intermingled with other code that makes use of the model. However, it is recommended that code for defining a model be clearly separated from code that makes use of the model or performs other functions. Specifically, we recommend a structure whereby a graph (or subgraph) is encapsulated in a function that creates the graph and returns the graph object, optionally along with some or all of the variables in the graph. Depending on the application, the graph creation function may take some application-specific arguments. For example:

```
function [fg, vars] = createMyGraph(xDim, yDim)
  fg = FactorGraph();
  v = Bit(xDim, yDim);
  p = Discrete(1:10);
  vars = {v, p};
  fg.addFactorVectorized(@exampleFactor, v(:, 1:end-1), v(:, 2:end), p);
  fg.addFactorVectorized(@exampleFactor, v(1:end-1, :), v(2:end, :), p);
end
```

This function can then be used by calling, for example:

```
[fg,vars] = createMyGraph(4,4);
```

While it is possible to use Dimple to retrieve variables from the factor graph object itself, returning variables in the graph creation function allows them to be more easily managed and manipulated. The set of returned variables should include, at least, the variables that will subsequently be conditioned on input data and the variables that will be queried after performing inference. But since which variables will be used for these or other purposes may not be known ahead of time, it is often useful to simply return all variables. If more than one variable or variable array is to be returned, this could be done either by returning each as a separate return value, or combining them all into a single cell array, as shown in the above example.

In a nested graph, it is often preferable to use a structure like that shown above at each layer of nesting. In this way, a sub-graph creation function might be reused in more than one different outer graph.

In general, functions that create a Dimple model should not include operations that are related to performing inference. This includes choosing the solver, setting parameters that affect the solver behavior. There may in some cases be exceptions. For example, while conditioning variables on input data would normally not be considered part of the model, in some situations, it might be appropriate to consider this part of the model and to include it in the model creation code.

4.2 Performing Inference

Once a model has been created in Dimple, the user may then perform inference on that model. This typically involves first conditioning on input data, then performing the inference computation, and then querying the model to retrieve information such as beliefs (marginals), maximum *a posteriori* (MAP) value, or samples from the posterior distribution.

4.2.1 Choosing a Solver

To perform the inference computation, Dimple supports a variety of *solver* algorithms that the user can choose from. For each solver, there are a number of options and configuration parameters that the user may specify to customize the behavior of the solver.

At a high level, Dimple supports three categories of solver:

- Belief propagation (BP)
- Gibbs sampling
- Linear programming (LP)

The BP solvers further segment into solvers that are sum-product based—used primarily to compute marginals of individual variables in the model:

- SumProduct
- JunctionTree
- ParticleBP

and solvers that are min-sum based—used to compute the maximum *a posteriori* (MAP) value:

- MinSum
- JunctionTreeMAP

In either case, the result may be approximate, depending on the specific model and solver settings.

In the case of sum-product, the solvers further divide based on how continuous variables are dealt with (Real, Complex, and RealJoint variables). The SumProduct solver (which is the default solver) uses a Gaussian representation for messages passed to and from continuous variables, while the ParticleBP solver uses a particle representation[9]. In the current version of Dimple, the min-sum solvers support only discrete variables[10].

The two forms of Junction Tree solver are variants of BP that provide exact inference results by transforming a loopy graphical model into a non-loopy graph by joining factors and variables and then performing the sum-product or min-sum algorithm on the transformed model. This is only feasible for models that are "narrow" enough, i.e., ones in which a large number of variables would not have to be removed to eliminate any loops. The Junction Tree solvers currently support only discrete variables (and continuous variables that have been conditioned with fixed values). See section 5.6.8 for more details.

The Gibbs solver supports all variable types, and may be used to generate samples from the posterior distribution, or to determine marginals or approximate the maximum *a posteriori* value (see section 5.6.9).

The LP solver transforms a factor graph over discrete variables into an equivalent linear program then solves this linear program (see section 5.6.11). This solver is limited to factor graphs containing only discrete variables.

The Solver is a property of a factor graph. To set the Solver for a given graph, this property is set to the name of the solver. For example:

```
myGraph.Solver = 'Gibbs';
```

The solver name is case insensitive. The current set of valid solver names are:[11]

- SumProduct
- MinSum
- JunctionTree
- JunctionTreeMAP
- ParticleBP
- Gibbs
- LP

More detail on each of these solvers is provided in section 5.6.

If no solver is specified for a graph, Dimple will use the SumProduct solver by default.

[9] In the current version of Dimple, the ParticleBP solver does not support Complex or RealJoint variables.

[10] This restriction may be lifted in a future version.

[11] In Dimple versions 0.04 and earlier, there was a separate Gaussian solver, which implemented the Gaussian BP for continuous variables. In subsequent versions of Dimple, this functionality has been incorporated into the SumProduct solver. For backward compatibility, the solver may be set to 'Gaussian,' but will use the SumProduct solver in this case.

4.2.2 Conditioning on Input Data

In many cases, the model created in Dimple represents the prior, before conditioning on the data. In this case, then assuming inference on the posterior model is desired, then the user must condition the model on the data before performing inference.

There are two primary ways to condition on input data. In the first approach, the values actually measured are not included in the model, and instead the effect of the data is specified via a likelihood function for each variable that is indirectly influenced by the data. In the second approach, the variables that will be measured are included in the model, and the value of each is fixed to the actual measured data value.

4.2.2.1 Using a Likelihood Function as Input

Suppose a variable to be measured, y, depends only on another variable, x, and the conditional distribution $p(y|x)$ is known. Then conditioned on the measured value, $y = Y$, then the likelihood of x is given by $L(x) = p(y = Y|x)$. If our model includes the variable x, but does not include y, then we can indicate the effect of measuring $y = Y$ by specifying the likelihood function $L(x)$ as the "input" to the variable x using the Input property of the variable.

The particular form of the Input property depends on the type of variable. For a Discrete variable type, the Input property is a vector with length equal to the size of the variable domain. The values represent the (not necessarily normalized) value of the likelihood function for each element of the domain. For example,

```
v = Discrete(1:4);
v.Input = [1.2, 0.6, 0, 0.8];
```

Notice that values in the Input vector may be greater than one—the Input is assumed to be arbitrarily scaled. All values, however, must be non-negative.

For a Bit variable, the Input property is specified differently. In this case, the Input is set to a scalar that represents a normalized version of the likelihood of the value 1. That is,

$$\frac{L(x = 1)}{L(x = 0) + L(x = 1)}$$

For example,

```
b = Bit();
b.Input = 0.3;
```

In this case, the value must be between 0 and 1, inclusive.

For a Real variable, the Input property is expressed in the form of a FactorFunction object that can connect to exactly one Real variable. The list of available built-in FactorFunctions is given in section 4.1.4.6. Typically, an Input would use one of the standard distributions included in this list. In this case, it must be one in which all the parameters can be fixed to pre-defined constants. For the Gibbs and ParticleBP solvers, any such factor function may be used as an Input. For the SumProduct solver, however, only a Normal factor function may be used. Below is an example of setting the Input for a Real variable:

```
r = Real();
r.Input = FactorFunction(’Normal’, measuredMean, measurementPrecision);
```

See section 5.2.6.2.3 for a description of how to set the Input on RealJoint (or Complex) variables.

If the inputs are to be applied to an array of variables, this can generally be done all at once by setting the Input property of the entire array. For a Discrete variable array, the Input is set to an array with the same dimensions as the variable array (or subarray), but with an extra dimensions corresponding to the Input vector for each variable in the array. (If a dimension in the array is 1, that dimension is not included.) For example:

```
v = Discrete(1:4, 2, 1);
v.Input = [1.2, 0.6, 0, 0.8 ; 0.4, 0, 1.1, 0.9];
```

For an array of Bit variables, the Input is an array with the same dimensions as the Bit array. For example:

```
b = Bit(2,3);
b.Input = [0.3 0.7 0.1; 0.0 0.8 0.6];
```

In the current version of Dimple, Inputs on Real variable arrays must be set one at a time, or all set to a single common value[12].

4.2.2.2 Fixing a Variable to a Known Value

In some cases, the variable that will be measured is included in the model. In this case, once the value becomes known, the variable may be fixed to that specific value so that the remainder of the model becomes conditioned on that value. The FixedValue property is used to set a variable to a fixed value. For a single variable, the FixedValue is set to any value in the domain of that variable. For example:

```
v = Discrete(1:4);
v.FixedValue = 2;
```

[12] This restriction may be removed in a future version.

```
v = Discrete([1.2, 5.6, 2.7, 6.94]);
v.FixedValue = 5.6;
```

```
b = Bit();
b.FixedValue = 0;
```

```
r = Real([-pi, pi]);
r.FixedValue = 1.7;
```

For Discrete variables, the FixedValue property is currently limited to variables with numeric domains, though the domains need not include only integer values[13].

For arrays of variables, the FixedValue property may be set for the entire array by setting it to an array of values of with the same dimensions as the array (or subarray) being set. For example:

```
b = Bit(2,3);
b.FixedValue = [0 1 0; 1 1 0];
```

To see if a FixedValue has been set on a variable, you can use the hasFixedValue method. For a single variable this method returns a boolean value, and for an array of variables this method returns a boolean array.

Because the Input and FixedValue properties serve similar purposes, setting one of these overrides any previous use of the other. Setting the Input property removes any fixed value and setting the FixedValue property removes any input[14].

4.2.2.3 Using a Data Source in a Rolled-Up Graph

In a rolled-up graph, the Input property of a variable can be set using a data source. Detail on how to do this can be found in section 4.3

4.2.3 Choosing a Schedule

All of the Dimple solvers operate by successively performing the inference computation on each element in the graph. In the case of BP solvers, both variable and factor nodes must be updated, and the performance of the inference can depend strongly on the order that

[13] This limitation may be removed in a future version.

[14] For implementation reasons, setting the fixed value of a Discrete or Bit variable also sets the Input to a delta function—with the value 0 except in the position corresponding to the fixed value that had been set.

these updates occur. Similarly, for the Gibbs solver, while variables must be updated in an order that maintains the requirements of valid Gibbs sampling, performance may depend on the particular order chosen.

The order of updates in Dimple is called a “schedule.” The schedule may either be determined automatically using one of Dimple’s built-in “schedulers,” or the user may specify a custom schedule.

Each solver has a default scheduler, so if the user does not explicitly choose one, a reasonable choice is made automatically.

4.2.3.1 Built-in Schedulers

If no scheduler or custom schedule is specified, a default scheduler will be used. The default scheduler depends on the selected solver.

Another scheduler may be specified by setting the Scheduler property of a graph:

```
myGraph.Scheduler = ’ExampleScheduler’;
```

When setting the scheduler for a graph, the name of the scheduler is case sensitive.

For the Junction Tree solvers, a tree schedule will always be used, so there is no need to choose a different scheduler. For the remaining BP solvers (SumProduct, MinSum, and ParticleBP), the following schedulers are available. More detail on each of these schedulers is provided in section 5.1.2.2.

- TreeOrFloodingScheduler
- TreeOrSequentialScheduler
- FloodingScheduler
- SequentialScheduler
- RandomWithoutReplacementScheduler
- RandomWithReplacementScheduler

In a nested graph, for most of the schedulers listed above (except for the random schedulers), the schedule is applied hierarchically. In particular, a subgraph is treated as a factor in the nesting level that it appears. When that subgraph is updated, the schedule for the corresponding subgraph is run in its entirety, updating all factors and variables contained within according to its specified schedule.

It is possible for subgraphs to be designated to use a schedule different from that of its parent graph. This can be done by specifying either a scheduler or a custom schedule for the subgraph prior to adding it to the parent graph. For example:

```
SubGraph.Scheduler = 'SequentialScheduler';
ParentGraph.addFactor(SubGraph, boundaryVariables);
ParentGraph.Scheduler = 'FloodingScheduler';
```

For the TreeOrFloodingScheduler and the TreeOrSequentialScheduler, the choice of schedule is done independently in the outer graph and in each subgraph. In case that a subgraph is a tree, the tree scheduler will be applied when updating that subgraph even if the parent graph is loopy. This structure can improve the performance of belief propagation by ensuring that the effect of variables at the boundary of the subgraph fully propagates to all other variables in the subgraph on each iteration.

For the RandomWithoutReplacementScheduler and RandomWithReplacementScheduler, if these are applied to a graph or subgraph, the hierarchy of any lower nesting layers is ignored. That is, the subgraphs below are essentially flattened prior to schedule creation, and any schedulers or custom schedules specified in lower layers of the hierarchy are ignored.

Because of the differences in operation between the Gibbs solver and the BP based solvers, the Gibbs solver supports a distinct set of schedulers. For the Gibbs solver, the following schedulers are available. More detail on each of these schedulers is provided in section 5.1.2.2.

- GibbsSequentialScanScheduler
- GibbsRandomScanScheduler

Because of the nature of the Gibbs solver, the nested structure of a graph is ignored in creating the schedule. That is, the graph hierarchy is essentially flattened prior to schedule creation, and only the scheduler specified on the outermost graph is applied.

4.2.3.2 Custom Schedules

Dimple supports user defined custom schedules created with a list of nodes and/or edges. A custom schedule is specified using the Schedule method. Specifying a custom schedule overrides any scheduler that the graph would otherwise use.

The following code demonstrates this feature:

```
   eq = @(x,y) x == y;
   fg = FactorGraph();
   a = Bit();
   b = Bit();
   c = Bit();
   eq1 = fg.addFactor(eq,a,b);
   eq2 = fg.addFactor(eq,b,c);
```

```
%define schedule
% update b
% update eq1->a
% update eq2->c
% update a->eq1
% update c->eq2
% update eq1->b
% update eq2->b
schedule = {
    b,
    {eq1,a},
    {eq2,c},
    {a,eq1},
    {c,eq2},
    {eq1,b},
    {eq2,b}
    };

fg.Schedule = schedule;

%Set priors
a.Input = .6;
b.Input = .7;
c.Input = .8;

%Solve
fg.NumIterations = 1;
fg.solve();
```

Dimple also supports nesting custom schedules and nesting in general. The following example demonstrates specifying nested graphs in a schedule.

```
eq = @(x,y) x == y;
b = Bit(2,1);
nfg = FactorGraph(b);
nfg.addFactor(eq,b(1),b(2));
b = Bit(3,1);
fg = FactorGraph();
nf1 = fg.addFactor(nfg,b(1),b(2));
nf2 = fg.addFactor(nfg,b(2),b(3));


fg.Schedule = {b(1),nf1,b(2),nf2,b(3)};
b(1).Input = .7;
fg.NumIterations = 1;
fg.solve();
```

And finally we look at nesting a custom schedule:

```
%Now let's try nesting with a custom schedule on the nested graph.

%Create a graph to nest and give it a funny schedule
% nfg: eb(1) - f1 - ib - f2 - eb(2)
eb = Bit(2,1);
```

```
    ib = Bit();
    nfg = FactorGraph(eb);
    f1 = nfg.addFactor(eq,eb(1),ib);
    f2 = nfg.addFactor(eq,ib,eb(2));
    %Set an input and solve
    eb(1).Input = .8;

    nfg.NumIterations = 1;
    nfg.solve();

    %We expect the output to be equal to the input since the tree
    %scheduler passes the info along.
    assertElementsAlmostEqual(eb(2).Belief,eb(1).Input(1));

    %Now we create a schedule that will not propagate the info.
    nfg.Schedule = {ib,{f1,eb(1)},{f2,eb(2)},eb(1),eb(2),f1,f2};
    nfg.solve();

    assertElementsAlmostEqual(eb(2).Belief,.5);

    %Nest it and see if the schedule is preserved
    b = Bit(2,1);
    fg = FactorGraph();
    g = fg.addFactor(nfg,b);

    fg.Schedule = {b(1),b(2),g};

    b(1).Input = .8;
    fg.NumIterations = 1;
    fg.solve();
    assertElementsAlmostEqual(b(2).Belief,.5);
```

4.2.4 Running the Solver

Once a factor graph has been created and conditioned on any input data, inference may be performed on the graph by calling the solve method:

```
myGraph.solve();
```

The solve method performs all necessary computation, leaving the results available to be subsequently accessed. The behavior of the solve method is determined by the chosen schedule as well as by any solver-specific configuration parameters.

For example, for all of the BP solvers, the number of iterations can be set. By default, the number of iterations is 1, but for a loopy factor graph, generally multiple iterations should be performed. To set the number of iterations prior to solving, the BPOptions.iterations option may be set on the graph:

```
myGraph.setOption('BPOptions.iterations', 10);
myGraph.solve();
```

As a shortcut, the NumIterations property of the graph may be used to set the option:

```
myGraph.NumIterations = 10;
myGraph.solve();
```

For the Gibbs solver, the BPOptions.iterations option isn't used and will be ignored if set; other options specific to this solver should be used instead. For example, to set the number of Gibbs samples to run before stopping (assuming the solver has been set to 'Gibbs'):

```
myGraph.setOption('GibbsOptions.numSamples', 1000);
myGraph.solve();
```

Note that in most cases solver options can be set directly on the factor graph and the values will only be used if the applicable solver is in use for that graph. Option values will usually not take effect until the solver objects have been initialized, but this is done automatically when the solve() method is run. If you need to interact directly with the solver representation of the factor graph, you can access it using the Solver property .

In general, each solver has a series of custom options and methods that can be used to configure the behavior of the solver and query its state. A complete list of these can be found in section 5.6.

In some cases, it is useful to observe the intermediate behavior of the solver before it has completed. For the BP solvers, this can be done by using the solver-specific iterate method instead of the solve method. When called without any arguments, this results in running one iteration. An optional argument allows specifying the number of iterations to run. Successive calls to iterate do not reset the state of the solver, allowing it to be called multiple times in succession. However, before running iterate for the first time, the initialize method must be called in order to reset the state before beginning. For example, here we run one iteration at a time, displaying the belief of a particular variable after each iteration:

```
myGraph.initialize();
for i=1:numberOfIterations
  myGraph.Solver.iterate();
  disp(someVariable.Belief);
end
```

If instead we wanted to run 5 iterations at a time, the iterate call would be replaced with:

```
myGraph.Solver.iterate(5);
```

For the Gibbs solver, a similar method allows running one or a specified number of samples at a time, skipping initialization as well as any burn-in or random restarts. This is the sample method, which behaves the same as the iterate method.

4.2.4.1 Multithreading

Some solvers support multithreading. The following option can be used to turn on multithreading:

```
fg.setOption('SolverOptions.enableMultithreading', true);
```

By default, multithreading is turned off. Once multithreading is turned on, for large graphs or large factors, users can see acceleration up to N times where N is the number of cores in their machine.

4.2.5 Getting the Results of Inference

Once the solver has been run, the results of inference can be obtained from the elements of the graph. The kinds of results that may be desired vary with the application, and the kinds of results that are available depend on the particular solver and other factors.

One of the most common types of results are beliefs on individual variables in the graph. The belief of a variable is an estimate of the marginal distribution of that variable given the graph and any conditioning data.

When available, the belief is accessed via the Belief property of a variable:

```
b = myVariable.Belief;
```

The particular form of the Belief property depends on the type of variable, and in some cases on the solver. For a Discrete variable type, the Belief property is a vector with length equal to the size of the variable domain. The values represent the normalized value of the approximate marginal probability of each element of the domain. For a Bit variable, the Belief property is a single number that represents the marginal probability of the value 1.

For Real variables when using the SumProduct solver, the Belief is represented the mean and precision parameters of a Normal distribution and for RealJoint variables they are represented as a mean vector and covariance matrix of a multivariate Normal distribution. Using the ParticleBP solver, beliefs are available for Real variables, but a different interface is used to obtain the beliefs in a useful form. This is summarized in section 5.6.10.4. Beliefs for Real variables are not currently supported in the Gibbs solver.

Beliefs can be obtained directly from an array of variables using the Belief property of the array. For a Discrete variable array, the Belief is an array with the same dimensions as the variable array (or subarray), but with an extra dimensions corresponding to the Belief vector for each variable in the array. (If a dimension in the array is 1, that dimension is not included.) For an array of Bit variables, the Belief is an array with the same dimensions as the Bit array.

Another useful inference result returned by the Value property of a variable. This property returns a single value from the domain of the variable that corresponds to the maximum value of the belief:

```
v = myVariable.Value;
```

As with the Belief property, this can be used either on individual variables or on a variable array. Support for this property is currently limited to discrete variables.

When using the Gibbs solver, there are additional inference results that may be useful. For example, for a given real variable, you can request the best sample value that has been seen during inference.

```
bestValue = myVariable.Solver.getBestSample();
```

This is defined as the value of that variable associated with the sample over all variables that resulted in the most probably configuration observed given the graph and any conditioning data. Considering the graph as a potential function over the configuration space of all variables, this corresponds to the lowest energy configuration that had been observed.

By default, the Gibbs solver doesn't save all samples, but if so configured for a given variable (or all variables) prior to running the solver, the solver will save all samples, allowing the entire set of samples (post burn-in) to be obtained.

```
myVariable.saveAllSamples();
myGraph.solve();
allSamples = myVariable.Solver.getAllSamples();
```

There are a number of other useful results that can be obtained from the Gibbs solver, which are detailed in section 5.6.9.

It is also possible to retrieve beliefs from factors. The belief of a factor is defined as the joint belief over all joint states of the variables connected to that factor. In the current version of Dimple, this works only for factors connected exclusively to discrete variables. The beliefs can be extracted using one of two properties of a factor:

```
fb = myFactor.Belief;
```

Using the Belief property results in a compact representation of the belief that leaves out values corresponding to zero values of the factor.

It is also possible to call:

```
fb = myFactor.FullBelief;
```

The FullBelief property results in a multidimensional array of beliefs over the range of all possible states of the connected variables. The more compact representation may be needed where the full representation would result in a data structure too large to be practical. More information about the Belief and FullBelief properties can be found in sections 5.3.2.1.1 and 5.3.2.1.2, respectively.

4.2.6 Explicit Scheduling and Retrieving Message Values

Dimple supports the ability to retrieve and set messages as well as to explicitly update edges, factors and variables.

```
%OK, first we create a simple Factor Graph with a single xor connecting two
%variables.
fg = FactorGraph();
b = Bit(2,1);
f = fg.addFactor(@xorDelta,b);
%We can go ahead and set some inputs
b(1).Input = .8;
b(2).Input = .7;


%we can examine some edges
disp(f.Ports{1}.InputMsg);
disp(f.Ports{1}.OutputMsg);

%we can even set some edge messages
f.Ports{1}.InputMsg = [.6 .4];

%we can update a node
b(1).update();
b(2).update();

%or all the variables in a vector.
b.update();

%or a specific edge
b(1).updateEdge(f);

%but updating via portNum is quicker
b(1).updateEdge(1);

%of course, if we don’t know the portNum, we can get it
portNum = b(1).getPortNum(f);
b(1).updateEdge(portNum);

%We can do the same kind of stuff with factors
f.updateEdge(b(1));
f.updateEdge(f.getPortNum(b(2)));

%Let’s look at some messages again
b(1).Ports{1}.InputMsg
b(2).Ports{1}.InputMsg

%and some beliefs
```

```
b.Belief
```

4.3 Using Rolled Up Factor Graphs

4.3.1 Markov Model

In our first rolled up graph we build a simple Markov model describing an infinite stream of variables.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Markov Model
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Build nested graph
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Here we build a graph that connects two variables by an xor
%equation An xor equation with only two variables is the
%equivalent of an equals constraint.

in = Bit();
out = Bit();
ng = FactorGraph(in,out);
ng.addFactor(@xorDelta,in,out);


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%create rolled up graph.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Now we build a FactorGraph that creates an infinite chain of %variables.
bs = BitStream();
fg = FactorGraph();

%Passing variable streams as arguments to addFactor will result
%in a rolled up graph.  Passing in a slice of a variable stream
%specifies a relative offset for where the nested graph should
%be connected to the variable stream.
fs = fg.addFactor(ng,bs,bs.getSlice(2));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%create data source
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
data = repmat([.4 .6]',1,10);
ds = DoubleArrayDataSource(data);
dsink = DoubleArrayDataSink();
bs.DataSource = ds;
bs.DataSink = dsink;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%solve
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
fg.solve();

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%get Beliefs
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
while dsink.hasNext()
    dsink.getNext()
end
```

Let's talk about some of the aspects of rolled up graphs illustrated in this example in more detail:

4.3.1.1 Variable Streams and Slices

Variable Streams represent infinite streams of variables. When instantiating a VariableStream, users have to specify a domain. The following are examples of instantiating Discrete VariableStreams:

- DiscreteStream(DiscreteDomain({0,1,2}));
- DiscreteStream({0,1,2}); - equivalent to the previous example
- DiscreteStream({0,1}); - A stream of bits
- BitStream(); - Shorthand for the previous example.

Users can optionally create streams of vectors of variables. When doing so, users have to specify the dimensions of their variables

- DiscreteStream([0 1 2],N,1) - For a column vector
- DiscreteStream([0 1 2],1,N) - For a row vector
- DiscreteStream([0 1 2],M,N,P) - For a three dimensional tensor

When adding a repeated FactorGraph, users need to be able to specify how a nested graph connects to the variable streams. Slices can be used to indicate where in the stream a nested graph should connect. Slices are essentially references to a Variable Stream with a specified offset and increment. Slices have two main methods:

- hasNext() – Returns true if the next variable referenced by the slice can be retrieved.
- getNext() – Returns the next variable in the slice if it can be referenced. Otherwise throws an error.

Users typically shouldn't use these methods. Slices should primarily be used as arguments to addFactor.

First we need to instantiate a variable stream:

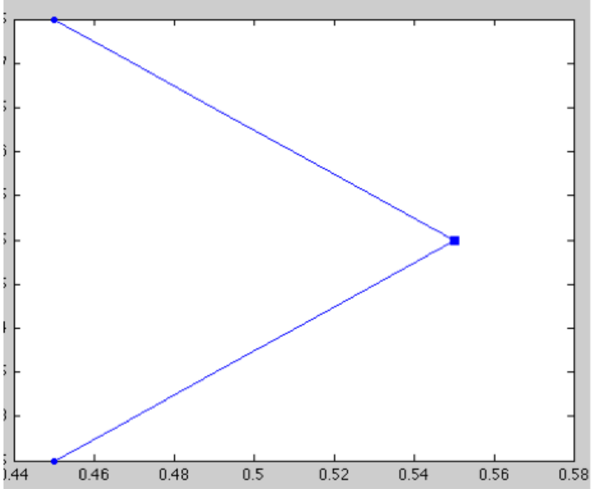
```
S = BitStream();
```

We can get slices in the following way:

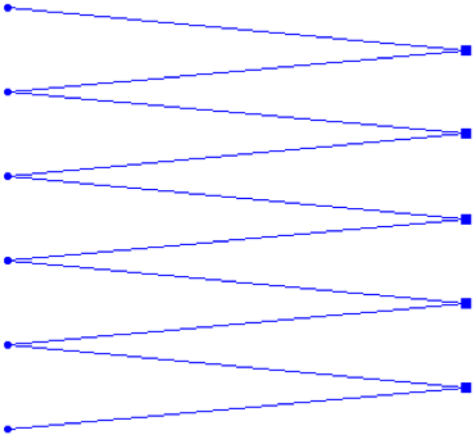
- S.getSlice(start); – Start specifies this slices starting point as an index into the variable stream.

4.3.1.2 Buffer Size

Users can specify a bufferSize for each FactorGraph stream. A FactorGraphStream is instantiated every time addFactor is called with a nestedGraph and VariableStreams or Slices. The default bufferSize is 1. Solving a graph with bufferSize one will result in a forward only algorithm. The bufferSize indicates how many nested graphs to instantiate for one step. In our Markov Model example, when buffer size is set to 1 and we plot the graph before solving we see this:



We see one factor and two instantiated variables. If we set bufferSize to 5 and plot we get:



We see five factors and 6 variables. After the first time we call 'advance' BlastFromThePast factors will be added to the oldest variable. These factors contain messages from the past. There are two ways to set the BufferSize for a FactorGraph stream:

- Fg.addFactor(ng,bufferSize,stream,slice,etc...); - Specified as second argument to addFactor.
- Fgs = fg.addFactor(ng,stream,slice,etc...); fgs.BufferSize = bufferSize; - Set on the FactorGraphStream directly.

4.3.1.3 DataSources

Users can create data sources and attach them to VariableStreams. As variables are created, data is retrieved from the DataSources and applied as inputs to the variables. If a VariableStream is never connected to a DataSource, hasNext() will always return true for that VariableStream. When a VariableStream is connected to a data source, hasNext() only returns true when there's more data in the DataSource.

DataSources implement the hasNext and getNext methods. Methods include:

- DoubleArrayDataSource(); – Constructor that creates an empty data source
- DoubleArrayDataSource(data); – Constructor that expects a matrix as input. The data should be formatted such that each column provides input data for a step in the rolled up graph.
- DoubleArrayDataSource(dimensions); – Dimensions is a row vector specifying the dimensions of the variable for each step of the repeated graph.
- DoubleArrayDataSource(dimensions,data); – Dimensions means the same as above. The data is added to the double array data source as though add(data) were called. See the following:

- add(data); – Users can add more data to the end of the data source. Data should have the following dimensions: [VarDimensions SizeOfInputData NumSteps). So if the user creates an MxN variable and wants to repeat it P times and the variable has domain of length K, the data should have the dimensions: [M N K P]

Dimple also has support for MultivariateDataSources.

- MultivariateDataSource() – Creates a data source object and assumes there is a single variable per step.
- MultivariateDataSource(dimensions) – Creates a data source objects that can be associated with a variable stream with the given dimensions.
- add(means,covariance) – Means should be in the form [VarDimensions NumberMeansPerVar] and Covariance should be of the form [VarDimensions NumberMeansPerVar NumberMeansPerVar]

4.3.1.4 DataSink

Users can retrieve their data using data sinks.

- DoubleArrayDataSink(); – Create a double array data sink
- DoubleArrayDataSink(dimensions) – Creates a double array data sink for a variable with the specified dimensions.
- MultivariateDataSink() – Created a multivariate data sink
- MultivariateDataSink(dimensions) – Creates a multivariate data sink with the specified dimensions.
- hasNext() – Is there more data in the data sink?
- getNext() – Retrieve the next chunk of data. For DoubleArrays, this returns data in the same form as is supplied to data sources. The MultivariateDataSink object returns MultivariateNormalParameters objects.
- varStream.DataSink = dataSink; – Data sinks can be assigned to variable streams.

4.3.1.5 Accessing Variables

In the absence of data sinks, users need a way to retrieve variables to get beliefs. The following methods allow the user to do that:

```
Vs = BitStream();
```

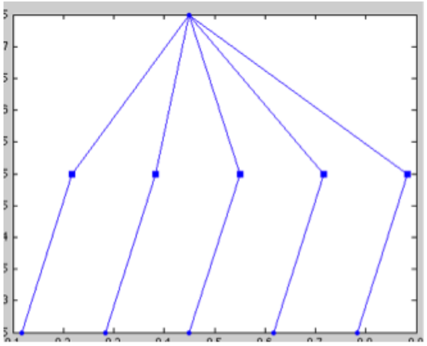
- Vs.Size – Number of variables in the buffer.
- Vs.get(index) – Retrieves a variable of the specified index. This is a 1-based value.

4.3.2 Markov Model with Parameter

When adding a repeated graph, it is possible to specify some variables as streams and others as individual variables. We sometimes call these individual variables parameters. Using this feature is straightforward:

```
ng = FactorGraph(a,b);
ng.addFactor(@xorDelta,a,b);
p = Bit();
s = BitStream();
fg = FactorGraph();
fgs = fg.addFactor(ng,p,s);
fgs.BufferSize = 5;
fg.plot();
```

This code results in the following graph:



4.3.3 Real Variables

Rolled up graphs work with real variables as well. Here we create another Markov Model. We use the SumProduct solver and a built-in 'Product' factor. We create a data source that only has information about the first variable. The means beliefs are growing by 110% as we iterate through the stream because the factor provides a constraint that each variable is 110% of the previous variable.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Build graph
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
a = Real();
b = Real();

ng = FactorGraph(a,b);
```

```
ng.addFactor('Product',b,a,1.1);

fg = FactorGraph();
s = RealStream();

fg.addFactor(ng,s,s.getSlice(2));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%set data
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

data = [[1; .1] repmat([0 Inf]',1,10)];
dataSource = DoubleArrayDataSource(data);


s.DataSource = dataSource;
s.DataSink = DoubleArrayDataSink();

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Solve
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
fg.solve();


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%get belief
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
while s.DataSink.hasNext();
    disp(s.DataSink.getNext());
end
```

This produces the following output:

```
    1.0000    0.1000

    1.1000    0.1100

    1.2100    0.1210

    1.3310    0.1331

    1.4641    0.1464

    1.6105    0.1611

    1.7716    0.1772

    1.9487    0.1949

    2.1436    0.2144
```

4.3.4 Manually Advancing

By default, rolled up graphs will advance until there is no data left in a DataSource. Users may override this behavior by either using FactorGraph.solveOneStep or setting the FactorGraph.NumSteps parameter. By default, NumSteps is set to Inf. By setting this to a finite number, N, users can examine the graph at every N steps of the rolled up graph. This allows the user to pull Beliefs off of any portion of the graph.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%simple Markov Model with larger buffer size

%Create the data
N = 10;
data = repmat([.4 .6]',1,N);

%Create a data source
dataSource = DoubleArrayDataSource(data);

%Create a variable stream.
vs = DiscreteStream({0,1});
vs.DataSource = dataSource;

%Create our nested graph
in = Bit();
out = Bit();
ng = FactorGraph(in,out);
ng.addFactor(@(a,b) a==b,in,out);

%Create our main factor graph
fg = FactorGraph();

%Build the repeated graph
bufferSize = 2;
fgs = fg.addFactor(ng, bufferSize, vs,vs.getSlice(2));

%Initialize our messages
fg.initialize();

while 1
    %Solve the current time step
    fg.solveOneStep();

    %Get the belief for the first variable
    belief = vs.get(3).Belief;
    disp(belief)

    if fg.hasNext
        fg.advance();
    else
        break;
    end
end
```

In this code snippet, the user initializes the graph and calls advance until there is no data left. At each step, the user retrieves the beliefs from the third instance of the variable in the variable stream.

The user can also progress N steps:

```
%Initialize our messages
fg.initialize();
fg.NumSteps = 2;

while 1
    %Solve the current time step
    fg.continueSolve(); %This method is need to avoid initialization

    %Get the belief for the first variable
    belief = vs.get(3).Belief;
    disp(belief)

    if fg.hasNext
        fg.advance();
    else
        break;
    end
end
```

4.4 Using Finite Field Variables

4.4.1 Overview

Dimple supports a special variable type called a FiniteFieldVariable and a few custom factors for these variables. They represent finite fields with $N = 2^n$ elements. These fields find frequent use in error correcting codes. Because Dimple can describe any discrete distribution, it is possible to handle finite fields simply by describing their factor tables. However, the native FiniteFieldVariable type is much more efficient. In particular, variable addition and multiplication, which naively require $\mathcal{O}(N^3)$ operations, are calculated in only $\mathcal{O}(N \log N)$ operations.

4.4.2 Finite Fields Without Optimizations

As we mentioned previously, a user can construct (non-optimized) finite fields from scratch.

First we create a domain for our variables using the MATLAB gf function (for Galois Field).

```
m = 3;
numElements = 2^m;
domain = 0:numElements-1;

tmp = gf(domain,m);
real_domain = cell(length(tmp),1);
for i = 1:length(tmp)
  real_domain{i} = tmp(i);
end
```

Next we create a bunch of variables with that domain.

```
x_slow = Discrete(real_domain);
y_slow = Discrete(real_domain);
z_slow = Discrete(real_domain);
```

Now we create our graph and add the addition constraint.

```
fg_slow = FactorGraph();

addDelta = @(x,y,z) x+y==z;
fg_slow.addFactor(addDelta,x_slow,y_slow,z_slow);
```

This code runs in $\mathcal{O}(N^3)$ time since it tries all combinations of x,y, and z. Next we set some inputs.

```
x_input = rand(size(x_slow.Domain.Elements));
y_input = rand(size(y_slow.Domain.Elements));

x_slow.Input = x_input;
y_slow.Input = y_input;
```

Finally we set number of iterations, solve, and look at beliefs.

```
fg_slow.NumIterations = 1;

fg_slow.solve();

z_slow.Belief
```

The solver runs in $\mathcal{O}(N^2)$ time since z is determined by x and y, x is determined by z, and y, and y is determined by x and z.

4.4.3 Optimized Finite Field Operations

Rather than building finite field elements from scratch, a user can use a built-in variable type and associated set of function nodes. These native variables are much faster, both for programming and algorithmic reasons. All of these operations are supported with the SumProduct solver.

4.4.3.1 FiniteFieldVariables

Dimple supports a FiniteFieldVariable variable type, which takes a primitive polynomial (to be discussed later) and dimensions of the matrix as constructor arguments:

```
v = FiniteFieldVariable(prim_poly,3,2);
```

This would create a 3x2 matrix of finite field Variables with the given primitive polynomial.

4.4.3.2 Addition

Users can use the following syntax to create an addition factor node with three variables:

```
myFactorGraph.addFactor('FiniteFieldAdd',x,y,z);
```

using the 'FiniteFieldAdd' built-in factor.

Adding this variable take $\mathcal{O}(1)$ time and solving takes $\mathcal{O}(N \log N)$ time, where N is the size of the finite field domain.

4.4.3.3 Multiplication

Similarly, the following syntax can be used to create a factor node with three variables for multiplication:

```
myFactorGraph.addFactor('FiniteFieldMult',x,y,z);
```

Under the hood this will create one of two custom factors, CustomFiniteFieldConstMult or CustomFiniteFieldMult. The former will be created if x or y is a constant and the latter will be created if neither is a constant. This allows Dimple to optimize belief propagation so that it runs in O(N) for multiplication by constants and O(Nlog(N)) in the more general case.

4.4.3.4 NVarFiniteFieldPlus

Suppose we have the finite field equation $x1 + x2 + x3 + x4 = 0$.

We can not express that using the FiniteFieldAdd factor directly, since it accepts only three arguments. However, we can support larger addition constraints by building a tree of these constraints. We do so using the following function:

```
NumVars = 4;
[graph,vars] = getNVarFiniteFieldPlus(prim_poly,NumVars);
```

This function takes a primitive polynomial and the number of variables involved in the constraint and builds up a graph such that $x_1 + ... + x_n = 0$.

It returns both the graph and the external variables of the graph. This can be used in one of two ways: setting inputs on the variables and solving the graph directly or using this as a nested sub-graph.

4.4.3.5 Projection

Elements of a finite field with base 2 can be represented as polynomials with binary coefficients. Polynomials with binary coefficients can be represented as strings of bits. For instance, $x^3 + x + 1$ could be represented in binary as 1011. Furthermore, that number can be represented by the (decimal) integer 11. When using finite fields for decoding, we are often taking bit strings and re-interpreting these as strings of finite field elements. We can use the FiniteFieldProjection built-in factor to relate n bits to a finite field variable with a domain of size 2^n.

The following code shows how to do that:

```
args = cell(n*2,1);
for j = 0:n-1
   args{j*2+1} = j;
   args{j*2+2} = bits(n-j);
end

myFactorGraph.addFactor('FiniteFieldProjection',v,args{:});
```

4.4.4 Primitive Polynomials

See Wikipedia for a definition.

4.4.5 Algorithmics

Dimple interprets the domains as integers mapping to bit strings describing the coefficients of polynomials. Internally, the FiniteFieldVariable contains functions to map from this representation to a representation of powers of the primitive polynomial. This operation is known as the discrete log. Similarly Dimple FiniteFieldVariables provide a function to map the powers back to the original representation (i.e., an exponentiation operator).

- The addition code computes $x+y$ by performing a fast Hadamard transform of the distribution of both x and y, pointwise multiplying the transforms, and then performing an inverse fast Hadamard transform.
- The generic multiplication code computes xy by performing a fast Fourier transform on the distribution of the non-zero elements of the distribution, pointwise multiplying the transforms, performing an inverse fast Fourier transform, and then accounting for the zero elements.
- The constant multiplication code computes x by converting the distribution of the non-zero values of x to the discrete log domain (which corresponds to reshuffling the array), adding the discrete log of modulo $N-1$ (cyclically shifting the array), and exponentiating (unshuffling the array back to the original representation).

4.5 Parameter Learning

Dimple currently has two supported parameter learning algorithms: Expectation-Maximization on directed graphs and PseudoLikelihood Parameter Estimation on undirected graphs. Both of these algorithms are provided as early stage implementations and the APIs will likely change in the next version of Dimple.

4.5.1 PseudoLikelihood Parameter Estimation on Undirected Graphs

The PseudoLikelihood Parameter Estimation uses the following as its objective function:

$$\ell_{PL}(\theta) = \frac{1}{M}\sum_{m}\sum_{i}\sum_{a\sim i}(x_{-i}^{(m)}, x_i^{(m)}) - \frac{1}{M}\sum_{m}\sum_{i} logZ(x_{N(i)};\theta)$$

Currently it uses a very naive gradient descent optimizer. Future versions will likely have pluggable optimizers for each learning algorithm. (Likely including algorithms like BFGS).

4.5.1.1 Creating a parameter learner

The following creates a learner and initializes a few variables:

```
pl = PLLearner(factorGraph,factorTables,variables);
```

Arguments:

- factorGraph - the Factor Graph of interest
- factorTables - a cell array of factor tables for which to learn parameters.
- variables - a cell array of variable matrices (the order must match your data ordering).

4.5.1.2 Learning

The following method runs pseudo likelihood gradient descent. After it is run, the factor tables will contain the values of the learned parameters. For now the optimizer is simply a routine that multiplies the gradient by a scale factor and applies that change to the parameters. In the future, optimizers will be first class citizens and can be plugged into learners.

```
args.numSteps = 100;
args.scaleFactor = 0.05;
pl.learn(samples,args);
```

Arguments:

- samples - An MxN matrix where M is the number of samples and N is the number of variables. Variable data must be specified in the same order the variables were specified in the learner's constructor. For now, this data specifies the domain indices, not the domain values. This should be fixed in the future (so the user can do either). In reality, we'll probably split out training data into a more interesting data structure. (Same with the optimizer)
- args.numSteps - How many gradient descent steps should the optimizer run.
- args.scaleFactor - The value by which we multiply the gradient before adding to the current parameters. oldParams = oldParams + scaleFactor*gradient

4.5.1.3 Batch Mode

Users can divide their samples into subsets to run pseudo likelihood parameter learning in "batch" mode. Assuming users have their samples stored in a cell array of matrices, they could iterate over the cell array as follows:

```
for i = 1:length(samples)
    pl.learn(samples{i},args);
end
```

4.5.1.4 Setting Data

When calling the learn routine, users can set the data. However, if users want some visibility into the gradient or the numerical gradient, they must first set the data using the setData method

```
pl.setData(samples)
```

Arguments:

- samples - Takes the same form as in the learn method.

4.5.1.5 Calculating the Pseudo Likelihood

Users can retrieve the pseudo likelihood given the currently set samples using the following code:

```
likelihood = pl.calculatePseudoLikelihood();
```

Return value:

- likelihood - The log pseudo likelihood.

4.5.1.6 Calculating the Gradient

For debugging purposes, the user can retrieve the gradient given the current sample set and parameter settings.

```
result = pl.calculateGradient()
```

Return values:

- result - MxN matrix where M is the number of factor tables being learned and N is the number of weights per factor table.

4.5.1.7 Calculating the Numerical Gradient

For debugging purposes, the user can return a numerical gradient

```
pl.calculateNumericalGradient(table, weightIndex, delta)
```

Arguments:

- table - Which table to modify
- weightIndex - Which weight index to modify
- delta - the delta (in the log domain) of the parameter.

4.5.2 Expectation-Maximization on Directed Graphs

See the FactorGraph.baumWelch method in the API section. see section 5.1.3.11

4.6 Graph Libraries

Dimple provides a few graphs that are useful as nested graphs.

4.6.1 Multiplexer CPDs

Suppose you wanted a factor representing a DAG with the following probability distribution:

$$p(Y = y|a, z_1, z_2, ...) \propto \delta(y = z_a)$$

where all variables are Discrete.

You could code this up in Dimple as follows:

```
function weight = myFunc(y,a,z)
    weight = y == z(a);
end

N = 3; %Number of possible sources
M = 2; %Domain of Zs and Y

y = Discrete(1:M);
a = Discrete(1:N);
z = Discrete(1:M,N,1);

fg = FactorGraph();

fg.addFactor(@myFunc,y,a,z);
```

However, to build this FactorTable takes $O(NM^{N+1})$ where N is the number of Zs and M is the domain size of the Zs. Runtime is almost as bad at $O(NM^N)$. However, there is an optimization that can result in $O(MN^2)$ runtime and graph building time. Dimple provides a MultiplexerCPD graph that can be used as a nested graph to achieve this optimization.

```
cpd = MultiplexerCPD({1,2},3);
Y = Discrete(1:2);
A = Discrete(1:3);
Z1 = Discrete(1:2);
Z2 = Discrete(1:2);
Z3 = Discrete(1:2);
fg = new FactorGraph();
fg.addFactor(cpd,Y,A,Z1,Z2,Z3);
```

Dimple supports each Z having different domains. In this case, Y's domain must be the sorted union of all the Z domains

```
cpd = MultiplexerCPD({{1,2},{1,2,3},{2,4}});
Y = Discrete(1:4);
A = Discrete(1:3);
Z1 = Discrete(1:2);
Z2 = Discrete(1:3);
Z3 = Discrete([2 4]);
fg = FactorGraph();
fg.addFactor(cpd,Y,A,Z1,Z2,Z3);
```

Note that when using the SumProduct solver, a custom implementation of the built-in ‘Multiplexer’ factor function (see section 5.9) exists that is even more efficient than using the MultiplexerCPD graph library function. When using other solvers with discrete variables, the MultiplexerCPD graph library function should be more efficient. When the Z variables are real rather than discrete, the ‘Multiplexer’ factor function is the only option.

4.6.2 N-Bit Xor Definition

An N-bit soft xor can be decomposed into a tree of three bit soft xors. Dimple provides the getNBitXorDef function to generate such a graph. The following code shows how to use such a graph.

```
fg = FactorGraph();
fg.addFactor(getNBitXorDef(4),Bit(4,1));
```

5 API Reference

The following section describes the functions, classes, properties, options and methods that comprise the Dimple API for MATLAB. In some cases, API documentation may also be obtained using either the MATLAB 'help' or 'doc' command with the name of the element of interest.

5.1 FactorGraph

The FactorGraph class represents a single factor graph and contains a collection of all factors and variables associated with that factor graph.

5.1.1 Constructor

```
FactorGraph([boundaryVariables])
```

For a basic factor graph, the constructor can be called without arguments.

For a nested factor graph (one that may be used as a sub-graph within another graph), the constructor must include a list of the boundary variables of the graph. When used as a sub-graph, the boundary variables are dummy variables with the same specification as the variables in the outer graph that will ultimately connect to the sub-graph. A graph defined with boundary variables may alternatively be used as a top-level graph, in which case the boundary variables are used directly.

5.1.2 Properties

5.1.2.1 Solver

Read-write. Indicates the choice of solver to be used for performing inference on the graph. The default solver is SumProduct.

When setting the solver, the solver is given by a string representing the name of the solver. The solver name is case insensitive.

```
fg.Solver = 'SolverName';
```

The current set of valid solver names are:

- SumProduct
- MinSum
- JunctionTree
- JunctionTreeMAP
- ParticleBP
- Gibbs

- LP

A description of each of these solvers is given in section 5.6.

Note that the solver can be modified at any time. After running the solver on a graph, the solver may be modified and the new solver run using the same graph[15].

5.1.2.2 Scheduler

Read-write. Indicates the scheduler to be used for performing inference on the graph (unless a custom schedule is specified instead). A scheduler defines a rule that determines the update schedule of a factor graph when performing inference.

When setting the scheduler, the scheduler is given by a string representing the name of the scheduler. The scheduler name is case *sensitive*.

```
fg.Scheduler = 'SchedulerName';
```

Each scheduler is applicable only to a certain subset of solvers. The list of all available built-in schedulers and a description of their behavior can be found in section 5.5.

5.1.2.3 Schedule

Read-write. Specifies a custom schedule to be used for performing inference. A custom schedule is in the form of a list of nodes or edges in the graph to be updated. Specifically, a cell array where each entry is either a node in the graph (either variable or factor), or a cell array containing a neighboring pair of nodes ({variable, factor} or {factor, variable}). The order of the entries in the cell array indicate the order that updates should be performed in a single iteration (or scan) when performing inference. Examples of using custom schedules are given in section 4.2.3.2.

For BP solvers, any of these entries may be included, and have the following interpretation[16].

Variable Update messages for all outgoing edges of that variable.

Factor Update messages for all outgoing edges of that factor.

{Variable, Factor} Update a single outgoing edge of the variable in the direction connecting to the specified factor.

[15] In this case, care must be taken to set any solver-specific parameters to the new values after changing the solver.

[16] When using the JunctionTree or JunctionTreeMAP solvers, the specified Schedule is ignored.

{Factor, Variable} Update a single outgoing edge of the factor in the direction connecting to the specified variable.

For BP solvers, a check is made when a custom schedule is set to ensure that all edges in the graph are updated at least once.

For the Gibbs solvers, the Schedule should include only variable entries. Any other entries will be ignored.

If a custom schedule is set on a factor graph (either an entire graph or a sub-graph), this schedule is used instead of any built-in scheduler that may have previously been set (or the default scheduler).

In a nested graph, the Schedule property at each nesting level may be set independently. For some built-in schedulers, the user may mix custom schedules at some nesting layers, while using built-in schedulers at others. The particular built-in schedulers that support such mixing are described in section 5.1.2.2.

5.1.2.4 NumIterations

Read-write. The NumIterations property sets the number of iterations BP will to run when using the solve method. This only applies to solvers that use BP, which are the SumProduct, MinSum, and ParticleBP solvers.

The default value is 1. For a factor graph with a tree-structure, when using the default scheduler, one iteration is appropriate. Otherwise, it would normally be appropriate to set the number of iterations to a larger value.

5.1.2.5 NumSteps

Read-write. This property is used for rolled-up graphs (see section 4.3). This property determines the number of steps over which to perform inference when using the solve or continueSolve methods (see sections 5.1.3.6 and 5.1.3.7). A *step* corresponds to a single run of the solver over the current portion of the rolled-up graph, followed by advancing the graph to the next position. By default, this property is infinite, resulting in no limit to the number of steps to be run (running until there is no more source data).

5.1.2.6 Name

Read-write. When read, retrieves the current name of the factor graph. When set, modifies the name of the factor graph to the corresponding value. The value set must be a string.

```
fg.Name = 'string';
```

5.1.2.7 Label

Read-write. All variables and factors in a Factor Graph must have unique names. However, sometimes it is desirable to have variables or factors share similar strings when being plotted or printed. Users can set the Label property to set the name for display. If the Label is not set, the Name will be used for display. Once the label is set, the label will be used for display.

5.1.2.8 Score

Read-only. When read, computes and returns the score (energy) of the graph given a specified value for each of the variables in the graph. The score represents the energy of the graph given the specified variable configuration, including all factors as well as all Inputs to variables (which behave as single-edge factors). The score value is relative, and may be arbitrarily normalized by an additive constant. The value of the score corresponds to the sum over factors and variables of their corresponding scores (see sections 5.3.1.1.4 and 5.2.2.1.6).

The value of each variable used when computing the Score is the Guess value for that variable (see section 5.2.2.1.5). If no Guess had yet been specified for a given variable, the value with the most likely belief (which corresponds to the Value property of the variable) is used[17].

For a rolled-up graph, the Score property represents only the score for only the portion of the graph in the current buffer.

5.1.2.9 BetheFreeEnergy

Read-only. (Only applies to the Sum Product Solver). When read returns:

$$BetheFreeEnergy = InternalEnergy - BetheEntropy$$

5.1.2.10 Internal Energy

Read-only. (Only applies to the Sum Product Solver). When read returns:

[17] For some solvers, beliefs are not supported for all variable types; in such cases there is no default value, so a Guess must be specified.

$$InternalEnergy = \sum_{a \in F} InternalEnergy(a) + \sum_{i \in V} InternalEnergy(i)$$

Where F is the set of all Factors and V is the set of all variables. If Dimple treated inputs as single node Factors, this method would only sum over factors.

For a definition of a Factor's InternalEnergy, see sections 5.3.1.1.5. For a definition of a Variable's InternalEnergy, see section 5.2.2.1.7.

5.1.2.11 Bethe Entropy

Read-only. (Only applies to the Sum Product Solver). When read returns:

$$BetheEntropy = \sum_{a \in F} BetheEntropy(a) - \sum_{i \in V} (d_i - 1) BetheEntropy(i)$$

Where F is the set of all Factors, V is the set of all variables, and d_i is the degree of variable i (i.e. the number of factors i is connected to).

For a definition of a Factor's BetheEntropy, see sections 5.3.1.1.6. For a definition of a Variable's InternalEnergy, see section 5.2.2.1.8.

5.1.3 Methods

5.1.3.1 addFactor

```
MyGraph.addFactor(factorSpecification, variableList);
```

The addFactor method is used to add a factor to a factor-graph, connecting that factor to a specified set of variables. There are several ways of specifying the factor. The method takes a factorSpecification argument followed by a comma-separated list of variables or variable arrays.

The list of variables or variable arrays indicates which variables to connect to the factor. The order of the variables listed must correspond to the order of edges of the factor. In the case of variable arrays, the array is flattened to a one-dimensional form using the usual MATLAB ordering, defining the order of the individual variables in the array. In this ordering, the array is scanned in successive dimensions, beginning with the first (row) dimension (equivalent to the MATLAB `(:)` operator).

Some of the variables may be replaced by constants. In this case, no variable is created, but instead the specified constant value is used in place of a variable for the corresponding edge of the factor. The value of a constant must be a valid value given the definition of the corresponding factor. Constants may be arrays, but in this case, they are not flattened. Instead they are treated as array-valued constants[18].

The factorSpecification may be specified in one of a number of different ways. The following table lists the various ways the factorSpecification can be specified:

Factor Specification	Description
functionHandle	The MATLAB function handle for a factor function. A factor function written in MATLAB is a function that takes arguments corresponding values from the domain of each of the connected variables (in the order listed) and returns a non-negative weight corresponding to the unnormalized value of the factor. In MATLAB, the function handle of a function is indicated using the `@` operator, for example, `@myFactor`. Anonymous functions are also supported. Specifying a factor in this form is supported only if all connected variables are discrete.
FactorTable	A FactorTable object as described in section 5.3.4. Specifying a factor in this form is supported only if all connected variables are discrete.
indexList, weightList	A factor table specified in an alternative form. The indexList and weightList arguments are defined identically to their definition in the FactorTable constructor described in section 5.3.4. Specifying a factor in this form is supported only if all connected variables are discrete.
weightArray	A factor table specified in an alternative form. The weightArray argument is defined identically to its definition in the FactorTable constructor described in section 5.3.4. Specifying a factor in this form is supported only if all connected variables are discrete.
builtInFactorName	String indicating the name of a built-in factor. The list of Dimple built-in factors is given in section 5.9. Referring to a built-in factor by name assumes no constructor arguments are needed for the corresponding FactorFunction. Built-in factor names are case sensitive.
FactorFunction	A FactorFunction object as described in section 5.3.3. This form can be used to specify a built-in factor that requires constructor arguments. The list of Dimple built-in factors is given in section 5.9.
factorFunctionAltForm	A factor function specified in an alternative form. The form is a cell array, where the first entry is a string indicating the name of the built-in factor, and the remaining entries are constructor arguments for the built-in factor.

[18] This is because Dimple supports variables that have domains for which individual domain elements are arrays.

Factor Specification	Description
FactorGraph	A sub-graph to be nested within this graph. The number and order of the variables listed in the variableList must correspond to the number and order of the boundary variables declared when the sub-graph was created. When adding a nested graph within an outer graph, the specified sub-graph is used as a template to create a new factor graph that is actually added to the outer graph. Copies are made of all of the variables and factors in the specified sub-graph.
javaFactorFunction	The fully-qualified name of a Java FactorFunction class. Creation of custom Java FactorFunctions is described in Appendix B.

5.1.3.2 addFactorVectorized

To get reasonable speed out of MATLAB, one needs to vectorize their code. If a user wishes to build a FactorGraph with large numbers of factors in a regular arrangement, they will want to avoid making many calls to addFactor. The addFactorVectorized method can be used instead to create many factors at once.

```
fg.addFactorVectorized(factorSpecification, variableList);
```

Using addFactorVectorized is very similar to using addFactor. The factorSpecification is defined identically to addFactor (see section 5.1.3.1). The variableList is similar, but includes some options for customizing how vectorization is done.

In the simplest case, the variableList includes a comma-separated list of variable arrays, all with the same dimensions. In this case, the result is creation of an array of factors with the same dimensions as the variables, where one element of each of the listed variable arrays is connected to the corresponding factor in the factor array.

One or more of the variables in the variableList may be a scalar (single variable) rather than an array. In this case, that variable is connected to all of the factors in the created factor array. The other variables in the variableList that are not scalars must all have the same dimensions.

In some cases, it may be useful to vectorize over some dimensions of a variable, but have other dimensions be flattened and connect to each individual factor in the factor array. In such cases, an entry in the variable list may be specified as a cell array containing two elements:

- The variable array
- A vector of indices, where each index represents a dimension over which the variable should be vectorized.

For example:

```
A = Bit(100,100);
B = Bit(100,100,5);
fg.addFactorVectorized(factorSpecification, A, {B, [1 2]});
```

For variable B, first two dimensions, of size 100x100 correspond to the 100x100 factors that we wish to vectorize factor creation. Since dimension 3 was not included in this list, then for each value of row and column dimension, the entire length of B variables in the third dimension are connected to the corresponding factor.

The vectorized dimensions for all non-scalar variables must be of the same size.

As in addFactor, some of the variables in variableList may be replaced with constants. In this case, every copy of the factor in the factor array uses the same constant value.

Further description and examples of using addFactorVectorized is given in section 4.1.4.2.

5.1.3.3 addFactorNoCache

When creating a factor for discrete variables, Dimple attempts to determine if the resulting factor tables would be the same as any previously created factors. If so, it shares the factor table to save space. In some cases, it may not be desirable for certain factors to share a factor table (for example if entries in one of the factor tables will later be modified). In such cases, the addFactorNoCache method can be used in lieu of the addFactor method. The interface for the addFactorNoCache is identical to that of the addFactor method.

5.1.3.4 addDirectedFactor

When adding a factor that will be designated as directed, this method provides a shorthand way to indicate this on factor creation.

```
fg.addDirectedFactor(factorSpecification, variableList, directedTo);
```

The factorSpecification is defined identically to addFactor (see section 5.1.3.1).

Instead of being a comma-separated list, the variableList argument is a cell array containing the list of variables or variable arrays to be connected to the factor. Otherwise, the behavior of the variable list is the same as for the addFactorMethod.

The directedTo argument is a cell array containing a subset of the variables listed in the variableList, specifically, the set of variables that are directed outputs of the factor.

Using addDirectedFactor is equivalent to the following (where the variableList in this case is a comma-separated list instead of a cell array):

```
f = fg.addFactor(factorSpecification, variableList);
f.directedTo(directedTo);
```

5.1.3.5 initialize

```
MyGraph.initialize();
```

The initialize method resets the state of the factor graph and its associated solver. When performing inference incrementally, for example using the iterate method, the initialize method must be called before the first iterate call. When using the solve method to perform inference, there is no need to call initialize first. The initialize method takes no arguments.

5.1.3.6 solve

```
MyGraph.solve();
```

The solve method runs the solver on the factor graph for the specified duration. Calling solve initializes the graph prior to solving.

For BP-based solvers, the solver runs the number of iterations specified by the NumIterations property. For the Gibbs solver, it runs for the specified number of samples (see section 5.6.9).

For rolled-up factor graphs, the solver runs the solver over multiple steps of the graph. A *step* corresponds to a single run of the solver over the current portion of the rolled-up graph, followed by advancing the graph to the next position. It performs the number of steps of inference specified by the NumSteps property or until there is no more data in a data source, whichever comes first.

5.1.3.7 continueSolve

This method is used for manually advancing a rolled-up graph (see section 4.3.4). This method takes no arguments and returns no value. When called, it performs the number of steps of inference specified by the NumSteps property or until there is no more data in a data source, whichever comes first. A *step* corresponds to a single run of the solver over the current portion of the rolled-up graph, followed by advancing the graph to the next position. The initialize method should be called prior to calling this method for the first time on an entire rolled-up graph, but should not be called before calling this method again to run additional steps.

5.1.3.8 solveOneStep

This method is used for manually advancing a rolled-up graph (see section 4.3.4). This method takes no arguments and returns no value. When called, it performs inference on the current portion of the rolled-up graph. Inference is performed on this section of the graph using whatever solver-specific parameters had previously been specified. The graph is not re-initialized prior to performing inference, starting instead from the final state resulting from inference on the previous graph position. The initialize method should be called prior to calling this method for the first time on an entire rolled-up graph, but should not be called after each advance to the next section.

5.1.3.9 advance

This method is used for manually advancing a rolled-up graph (see section 4.3.4). This method takes no arguments and returns no value. When called, it advances the graph to the next position. Advancing the graph involves getting the next value of all data sources, and writing the next available value to all data sinks.

5.1.3.10 hasNext

This method is used for manually advancing a rolled-up graph (see section 4.3.4). This method takes no arguments. It returns a boolean value indicating whether or not it is possible to advance the graph to the next position. This will be true only if all of the data sources have at least one more available value.

5.1.3.11 baumWelch

```
fg.baumWelch(factorList, numRestarts, numSteps);
```

The baumWelch method performs the Expectation-Maximization (EM) algorithm on a factor graph (the specific case of EM on an HMM graph is called the Baum-Welch algorithm, though EM in Dimple can be applied to any graph structure).

This method has the following limitations:

- The factors on which parameter estimation is being performed must be directed (see section 5.3.1.1.3).
- The factors on which parameter estimation is being performed must be connected only to discrete variables.
- The Solver must be set to use the SumProduct solver (the default solver).

The factorList argument is either a single Factor or FactorTable, or a cell array of Factors or FactorTables. The weight values in these factor tables are the parameters to be estimated. All other factors in the graph remain fixed and unmodified. When including a FactorTable, if that table is used in more than one factor in the graph, the entries in the table are tied, estimating them jointly for all factors containing that table.

The numRestarts argument indicates the number of times the EM process is repeated using distinct random restart values for the parameters. When numRestarts is greater than 1, the EM process is repeated with different random initialization, and the final resulting parameter values are chosen from the run that resulted in the smallest Bethe free energy for the graph.

The numSteps argument indicates (for each restart), how many steps of the EM algorithm are to be run. A single step of the EM algorithm is one run of belief propagation followed by re-estimation of the parameter values.

Note that the number of iterations for each run of belief propagation is determined from the NumIterations property. If the graph is a tree, the number of iterations should be 1 (the default value). If the graph is loopy, it should be set to a larger value (in this case, the EM algorithm is only approximate).

Upon completion of this method, the result appears in the factor tables that were listed in the factorList argument. That is, the factor table weights contain the resulting estimated values. To read these values, the Weights property of the factor table can be read (see section 5.3.4). For a factor, the factor table can be extracted using the FactorTable property (see section 5.3.4), and then the weights can be read from that.

5.1.3.12 join

The join method can be used to join a set of existing variables or a set of existing factors in a graph. In the current version of Dimple this method is supported only for discrete variables, and factors connected only to discrete variables.

When joining variables, the join method is called with a comma-separated list of variables to be joined.

```
fg.join(variableList);
```

The result is a new variable with a domain that is the Cartesian product of the domains of all of the variables being joined.

When joining factors, the join method is called with a comma-separated list of factors to be joined.

```
fg.join(factorList);
```

The result is a new factor with a factor table that corresponds to the product of the factors being joined. The new factor connects with the union of all variables that were previously connected to any of the joined variables.

5.1.3.13 split

The split method splits a variable in an existing graph into two variables connected by an Equality factor.

```
fg.split(variable, [factorList]);
```

The method takes an optional comma-separated list of factors. This list of factors identifies factors already connected to the variable that are to be moved to the new instance of the variable. All unspecified factors remain connected to the original instance.

5.1.3.14 removeFactor

```
fg.removeFactor(factor);
```

This method removes the specified factor from an existing factor graph that contains it. This also removes all edges that connect this factor to neighboring variables.

5.1.3.15 plot

The plot method is used to visualize a factor graph or a portion of a factor graph. Examples of various use cases for plotting factor graphs are given in section 4.1.6.

The plot method may be called with no arguments, or with a set of optional arguments. The optional arguments are defined in the following table. These are in the form of names (as strings) followed by one or more values.

Name	Values	Description
`'labels'`	true/false	Boolean value indicating whether or not to display the name of each variable and factor node. By default labels are not displayed.
`'color'`	[nodeList], color	Specifies the color of some or all nodes in the graph. If no nodeList is included, then this specifies the color of all nodes in the graph. If nodeList is specified, then the color applies to only the listed nodes. The nodeList argument is a cell-array of nodes in the graph (factors and variables). The color argument is a string indicating a MATLAB color (e.g., 'r'). More than one set of 'color' arguments may appear in a single call to plot.
`'nodes'`	nodeList	Indicates the set of nodes in the graph to be included in the plot. The nodeList argument is a cell-array of nodes in the graph (factors and variables). By default the entire graph is plotted.
`'depth'`	rootNode, depth	This argument indicates a portion of the graph to be displayed by specifying a root node and depth (distance) from that node. The plot includes only the root node and all other nodes that are within the specified distance from the root. By default the entire graph is plotted.
`'nesting'`	nestingDepth	By default, the plotting method ignores hierarchy and plots the flattened graph. The nesting argument specifies how deep to descend into the nesting hierarchy before considering nested graphs to be factors and plotting them as such.

5.1.3.16 addBoundaryVariables

```
fg.addBoundaryVariables([factorList]);
```

This method takes a comma separated list of variables. The listed variables can then be used as boundary variables in a nested graph.

```

ng = FactorGraph();
a = Bit(2,1);
y = a(1) + a(2);
ng.addBoundaryVariables(y,a);

fg = FactorGraph();
y = Discrete(0:2);
a = Bit(2,1);
fg.addFactor(ng,y,a);
a.Input = [1 0];
fg.solve();
```

5.1.4 Introspection

The FactorGraph class provides several feature for inspecting aspects of the graph. The ability to nest graphs complicates things a bit. Nested FactorGraphs can be considered Factors. All of the introspection features allow the user to view nested graphs as leaf factors or to descend into them and operate on the children of the nested graphs. Each feature provides several methods:

- <FeatureName>(int relativeNestingDepth) – The relativeNestingDepth specifies how deep to descend into the nested FactorGraphs before treating deeper NestedGraphs as Factors. Specifying 0 will treat the top level nested Graphs as factors. Specifying a large enough number will descend all the way to the leaf factors. Specifying something between 0 and the FactorGraph's maximum depth will descend as far as this parameter specifies before considering NestedGraphs to be factors. The parameter contains the word "relative" because users can retrieve nested graphs. They can call one of the feature's methods on that nested graph.
- <FeatureName>Flat() – equivalent of <FeatureName>(max int)
- <FeatureName>Top() – equivalent of <FeatureName>(0)
- <FeatureName>() – equivalent of <FeatureName>Flat(). It was thought that users will most often want to operate on the FactorGraph in its flattened form.

Now, on to the specific features.

5.1.4.1 Retrieving All Factors

Users can retrieve Factors and/or NestedGraphs associated with a graph using the Factors methods and properties:

- Fg.Factors
- Fg.FactorsFlat
- Fg.FactorsTop
- Fg.getFactors(relativeNestingDepth)

When the user specifies a relativeNestingDepth or calls FactorsTop, the resulting cell array will contain a mix of leaf factors and Nested Graphs.

5.1.4.2 Retrieving Factors but Not Nested Factor Graphs

The FactorGraph class provides the following:

- NonFactorGraphFactors
- NonFactorGraphFactorsFlat
- NonFactorGraphFactorsTop
- getNonFactorGraphFactors(relativeNestingDepth)

As the name implies, this will behave similar to the Factors properties and methods but will exclude nested graphs.

5.1.4.3 Retrieving Variables

The FactorGraph class provides the following:

- Variables – calls VariablesFlat
- VariablesFlat – Returns a list of all the Variables in the graph, including those contained by nested graphs.
- VariablesTop – Returns only those variables contained in the top level of the graph.
- getVariables(relativeNestingDepth,forceIncludeBoundaryVariables) – Returns all variables contained in the FactorGraph from which the method is called as Variables that are as deep as the specified relativeNestingDepth. The second parameter is optional and defaults to false. When false, boundary variables are only included by the root graph. When true, boundary variables are included regardless of whether a graph is a root or nested graph.

5.1.4.4 Retrieving All Nodes

The FactorGraph provides the following:

- Nodes
- NodesFlat
- NodesTop
- getNodes(relativeNestingDepth,forceIncludeBoundaryVariables)

These methods call the Factor and Variable methods and concatenate the results together.

5.1.4.5 Determining if a FactorGraph is a tree

The FactorGraph class provides the following:

- isTree(relativeNestingDepth)
- isTreeTop
- isTreeFlat

isTree – Users can call <factor graph name>.isTree() to determine if a FactorGraph is a tree. If the graph contains cycles, this method will return false. Like the other methods, the relativeNestingDepth determines at what point to consider NestedGraphs to be leaf nodes.

5.1.4.6 Retrieving an Adjacency Matrix

All of the following methods return a pair: [A, labels] where A is a square connectivity matrix and labels is a cell array of strings specifying the names of the nodes in A.

- getAdjacencyMatrix(relativeNestingDepth,forceIncludeBoundaryVariables) – relativeNestingDepth behaves the same as in other methods that take this parameter. So does forceIncludeBoundaryVariables. forceIncludeBoundaryVariables has a default value of false.
- getAdjacencyMatrix(nodes,forceIncludeBoundaryVariables) – Users can specify a specific subset of nodes in which they are interested. This method will return an adjacency matrix with only those nodes. Nodes are considered connected only if there is an edge directly connecting them.
- getAdjacencyMatrixTop() – equivalent to getAdjacencyMatrix(0,false)
- getAdjacencyMatrixFlat() – equivalent to getAdjacencyMatrix(intmax,false)

FactorGraph also provides an AdjacencyMatrix Property:

- AdjacencyMatrix – equivalent to getAdjacencyMatrixFlat and only returns A (not the labels). MATLAB properties can only return one object.

An example of getAdjacencyMatrix:

```
fg = FactorGraph();
b = Bit(2,1);
b(1).Name = 'b1';
```

```
b(2).Name = 'b2';
f = fg.addFactor(@xorDelta,b);
f.Name = 'f';
[A,labels] = fg.getAdjacencyMatrix();
A =

     0     0     1
     0     0     1
     1     1     0


labels =

    'b1'
    'b2'
    'f'
```

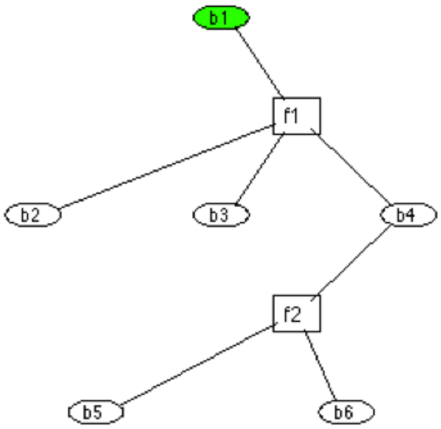
5.1.4.7 Depth First Search

- depthFirstSearch(node, searchDepth, relativeNestingDepth) –
  - node – Specifies the node from which to initiate the search
  - searchDepth – specifies how far from node the search should go.
  - relativeNestingDepth – determines how deep to go down the NestedGraphs before considering NestedGraphs to be leaf nodes.
- depthFirstSearchFlat(node, searchDepth) – equivalent of depthFirstSearch(node,searchDepth,maxint)
- depthFirstSearchThop(node, searchDepth) – equivalent of depthFirstSearch(node,searchDepth,0)

An example:

```
fg = FactorGraph();
b = Bit(6,1);
for i = 1:6
b(i).Name = sprintf('b%d',i);
end
f1 = fg.addFactor(@xorDelta,b(1:4));
f1.Name = 'f1';
f2 = fg.addFactor(@xorDelta,b(4:6));
f2.Name = 'f2';

nodes = fg.depthFirstSearch(b(1),3);
```

calling fg.plot('color',b(1),'g','labels',true) reveals the following structure of this graph



As you might guess fg.depthFirstSearch(b(1),3) will return a collection of six nodes: b1, f1, b2, b3, b4, and f2. It will not include b5 and b6 since those are at a depth of four from b1.

5.2 Variables and Related Classes

5.2.1 Variable Types

The following variable types are defined in Dimple. Some variable types are supported by only a subset of solvers. The following table lists the Dimple variable types and the solvers that support them.

Variable Type	Supported Solvers
Discrete	all
Bit	all
Real	SumProduct, Gibbs, ParticleBP
RealJoint	SumProduct, Gibbs
Complex	SumProduct, Gibbs
FiniteFieldVariable	all[19]

5.2.2 Common Properties and Methods

The following properties and methods are common to variables of all types.

5.2.2.1 Properties

5.2.2.1.1 Name

Read-write. When read, retrieves the current name of the variable or array of variables. When set, modifies the name of the variable to the corresponding value. The value set must be a string.

```
var.Name = ’string’;
```

When setting the Name, only one variable in an array may be set at a time. To set the names of an entire array of variables to distinct values, the setNames method may be used (see section 5.2.2.2.1).

5.2.2.1.2 Label

[19] The performance enhanced implementations of built-in factors for FiniteField variables are only available when using the SumProduct solver.

Read-write. All variables and factors in a Factor Graph must have unique names. However, sometimes it is desirable to have variables or factors share similar strings when being plotted or printed. Users can set the Label property to set the name for display. If the Label is not set, the Name will be used for display. Once the label is set, the label will be used for display.

```
var.Label = 'string';
```

5.2.2.1.3 Domain

Read-only. Returns the domain in a form that depends on the variable type, as summarized in the following table:

Variable Type	Domain Data Type
Discrete	DiscreteDomain (see section 5.2.9)
Bit	DiscreteDomain (see section 5.2.9)
Real	RealDomain (see section 5.2.10)
RealJoint	RealJointDomain (see section 5.2.11)
Complex	ComplexDomain (see section 5.2.12)
FiniteFieldVariable	FiniteFieldDomain (see section 5.2.13)

5.2.2.1.4 Solver

Read-only. Returns the solver-object associated with the variable, to which solver-specific methods can be called. See section 5.6, which describes the solvers, including the solver-specific methods for each solver.

5.2.2.1.5 Guess

Read-write. Specifies a value from the variable to be used when computing the Score of the factor graph (or of the variable or neighboring factors). The Guess must be a valid value from the domain of the variable.

If the Guess had not yet been set, its value defaults to the most likely belief (which corresponds to the Value property of the variable)[20].

5.2.2.1.6 Score

[20] For some solvers, beliefs are not supported for all variable types; in such cases there is no default value, so a Guess must be specified in order to compute the Score.

Read-only. When read, computes and returns the score (energy) of the Input to this variable, which is treated as a single-edge factors, given a specified value for the variable. The score value is relative, and may be arbitrarily normalized by an additive constant.

The value of the variable used when computing the Score is the Guess value for this variable (see section 5.2.2.1.5). If no Guess had yet been specified, the value with the most likely belief (which corresponds to the Value property of the variable) is used[21].

5.2.2.1.7 Internal Energy

Read-only. (Only applies to the Sum Product Solver). When read, returns:

$$InternalEnergy(i) = \sum_{d \in D} B_i(d) * (-log(Input(d)))$$

Read-only. When read returns:

Where D is variable i's domain, Input is the variable's input, and B_i is the variable Belief.

5.2.2.1.8 Bethe Entropy

Read-only. (Only applies to the Sum Product Solver). When read, returns:

$$BetheEntropy(i) = -\sum d \in DB_i(d) * log(B_i(d))$$

Where D is variable i's domain and B_i is the variable Belief.

5.2.2.1.9 Ports

Read-only. Retrieves a cell array containing a list of Ports connecting the variable to its neighboring factors.

5.2.2.2 Methods

[21] For some solvers, beliefs are not supported for all variable types; in such cases there is no default value, so a Guess must be specified.

5.2.2.2.1 setNames

For an array of variables, the setNames method sets the name of each variable in the array to a distinct value derived from the supplied string argument. When called with:

```
varArray.setName('baseName');
```

the resulting variable names are of the form: `baseName_vv0`, `baseName_vv1`, `baseName_vv2`, etc., where each variable's name is the concatenation of the base name with the suffix `_vv` followed by a unique number for each variable in the array.

5.2.2.2.2 invokeSolverSpecificMethod

```
variableArray.invokeSolverSpecificMethod('methodName', arguments);
```

For an array of variables, the invokeSolverSpecificMethod calls the specified solver-specific method on each of the variables in the array. Here, 'methodName' is a text string with the name of the solver-specific method, and arguments is an optional comma-separated list of arguments to that method. This method does not result in any return values. For solver specific methods that return results, use the invokeSolverSpecificMethodWithReturnValue method instead.

5.2.2.2.3 invokeSolverSpecificMethodWithReturnValue

```
returnArray = variableArray.invokeSolverSpecificMethodWithReturnValue('
    methodName', arguments);
```

For an array of variables, the invokeSolverSpecificMethodWithReturnValue calls the specified solver-specific method on each of the variables in the array, returning a return value for each variable. Here, 'methodName' is a text string with the name of the solver-specific method, and arguments is an optional comma-separated list of arguments to that method. The returnArray is a cell-array of the return values of the method, with dimensions equal to the dimensions of the variable array.

5.2.2.3 Operators

5.2.2.3.1 Operators for Implicit Factor Creation

Dimple supports a set of overloaded MATLAB operators and functions that operate on variables to implicitly add factors to a factor graph.

The list of supported operators and functions is given in section 5.10. A description of how to use these operators and functions is given in section 4.1.4.7.

Using one of the defined operators or functions with one or more variables will result in creation of a new variable, which can be assigned to a new variable name, with the appropriate domain. It will also result in the creation of a factor which is added to the most recently created factor graph. This factor will connect to the input variables and the newly created result variable. For example:

```
c = a + b;
```

This results in the creation of a Sum factor with connected variables, c, a, and b (in that order).

These operators and functions can be compounded into a single line of code, resulting in creation of intermediate anonymous variables. For example:

```
z = (a + b) * c^d - sqrt(-e);
```

Like using the addFactor method, some of the inputs to these operators and functions may be constants. Specifically, for binary operators, one of the inputs may be a constant instead of a variable. For example:

```
x = a^2;
y = (a + b + 2) * 3;
z = a * (2 + 1i);
```

These operators and functions can be applied to arrays of variables, with some limitations. Specifically, if each of the input variables are vectors of the same dimension, then the result will be to create a vector of output variables of the same dimension, along with a vector of factors relating the inputs and outputs.

In some cases, to be consistent with MATLAB notation, there is a distinction made between the vectorized and non-vectorized operator. Specifically, Dimple uses MATLAB's notation for pointwise product and power operators to indicate a vectorized operation. For example, if variables a through e are vectors of variables of identical size, then the following would create a variable vector z, and a series of factors relating these variables.

```
z = (a .* b) + c.^d - sqrt(-e);
```

For binary operators, one of the inputs may be a scalar variable or a scalar constant instead of a variable vector. For a scalar variable, the result is that scalar variable connecting to each instance of the factors that are created. For a constant, each instance of the factor uses the same constant for that input (vectors of distinct constants are not currently supported).

5.2.2.3.2 repmat

Dimple overloads the MATLAB repmat function when called with a Dimple variable or variable array as its first argument. The result of this function is an array of variables in the requested dimensions. Dimple does not actually make multiple copies of the variables, but instead creates a variable array that provides repeated references to the same variables in the existing array.

For example:

```
var = Bit(10, 1);
varRef = repmat(var, 1, 10);
```

In this case varRef is a 10x10 array of variables. Each element of varRef, however, is not a distinct Dimple variable, but a reference to an element in var, where for each row of varRef, there are 10 repeated copies of the corresponding variable in var. Use of repmat for variables can be useful when using addFactorVectorized (see sections 4.1.4.2 and 5.1.3.2).

5.2.3 Discrete

A Discrete variable represents a variable that can take on a finite set of distinct states. The Discrete class corresponds to either a single Discrete variable or a multidimensional array of Discrete variables. All properties/methods can either be called for all elements in the collection or for individual elements of the collection.

5.2.3.1 Constructor

The Discrete constructor can be used to create an N-dimensional collection of Dimple Discrete variables. The constructor is called with the following arguments (arguments in brackets are optional).

```
Discrete(domain, [dimensions])
```

- domain is a required argument indicating the domain of the variable. The domain may either be a numeric array of domain elements, a cell array of domain elements, or a DiscreteDomain object (see section 5.2.3.1.1).

- dimensions is an optional variable-length comma-separated list of matrix dimensions (an empty list indicates a single Discrete variable).

For example:

```
domain = [0 1 2];
w = Discrete(domain);
x = Discrete(domain, 4);
y = Discrete(domain, 2, 3);
z = Discrete(domain, 2, 3, 4);
```

We examine each of these arguments in more detail in the following sections.

5.2.3.1.1 Domain

Every Discrete random variable has a domain associated with it. A domain is a set. Elements of the set may be any object type. For example, the following are Discrete variables with valid domains:

```
a = Discrete({1, 2, 3});
b = Discrete({1+i, i, 2*i});
c = Discrete({[1 0; 0 1], [i 1, 2*i 1]});
d = Discrete({[1 0; 0 1], 2, i+1});
e = Discrete({1.2, 3, pi/2});
f = Discrete({'red', 'green', 'blue'});
```

(a) creates a variable whose domain consists of three values: 1, 2, and 3. (b) creates a variable whose domain consists of three complex numbers. (c) creates a variable whose domain consists of two elements, each of which is a 2x2 complex matrix. (d) creates a variable whose domain consists of three elements: a matrix, real scalar, and complex scalar. (e) creates a variable whose domain consists of both floating-point and integer values. (f) creates a variable whose domain is a set of strings.

In the previous example we used cell arrays to specify the elements of a domain. When the domain consists only of numeric values (integer or floating-point), domains can instead be specified as a numeric array. In this case, each element of the array (regardless of the array's dimensions) is considered a distinct entry in the domain.

```
a = Discrete(0:2);
b = Discrete([1 2 3; 4 5 6]);
c = Discrete([0:2]');
```

(a) creates a variable with a domain of 0, 1, and 2. (b) creates a variable with a domain of 1, 2, 3, 4, 5, 6. (c) creates a variable with domain of 0, 1, and 2.

The domain may also be specified using a DiscreteDomain object. In that case, the domain of the variable consists of the elements of this object. For example:

```
myDomain = DiscreteDomain(0:10);
a = Discrete(myDomain);
```

See section 5.2.9 for more information about the DiscreteDomain class.

5.2.3.1.2 List of Matrix Dimensions

If the variable constructor is called without any dimensions, a single variable will be created.

If one dimension n is specified, a square array of dimensions n x n variables will be created[22].

With k dimensions specified, n1, n2, ..., nk, a multidimensional variable array of dimensions n1 x n2 x ... x nk will be created.

5.2.3.2 Properties

5.2.3.2.1 Belief

Read-only. For any single variable, the Belief method returns a vector whose length is the total number of elements of the domain of the variable. When called after running a solver to perform inference on the graph, each element of the vector contains the estimated marginal probability of the corresponding element of the domain of the variable. The results are undefined if called prior to running a solver.

For an array of variables, the Belief method will return an array of vectors (that is, an array one dimension larger than the variable array) containing the beliefs of each variable in the array.

5.2.3.2.2 Value

Read-only. In some cases, one may wish to retrieve the single most likely element of a variable's domain. The Value property does just that.

For any single variable, the Value method returns a single value chosen from the domain of the variable. When called after running a solver to perform inference on the graph, the value returned corresponds to the element in the variable's domain that has the largest

[22] This follows a common MATLAB convention.

estimated marginal probability[23]. The results are undefined if called prior to running a solver.

For an array of variables, the Value method will return an array of values, each from the domain of the corresponding variable representing the largest estimated marginal probability for that variable.

5.2.3.2.3 Input

Read-write. For any variable, the Input method can be used to set and return the current input of that variable. An input behaves much like a single edge factor connected to the variable, and is typically used the represent the likelihood function associated with a measured value (see section 4.2.2.1).

When read, for a single variable returns an array of values with each value representing the current input setting for the corresponding element of the variable's domain. The length of this array is equal to the total number of elements of the domain. When read, for an array of variables, the result is an array with dimension one larger than the dimension of the variable array. The additional dimension represents the current set of input values for the corresponding variable in the array.

When written, for a single variable, the value must be an array of length equal to the domain of the variable. The values in the array must all be non-negative, and non-infinite, but are otherwise arbitrary. When written, for an array of variables, the values must be a multidimensional array where the first set of dimensions exactly match the dimensions of the array (or the portion of the array) being set, and length of the last dimension is the number of elements in the variable's domain.

5.2.3.2.4 FixedValue

Read-write. For any variable, the FixedValue property can be used to set the variable to a specific fixed value, and to retrieve the fixed-value if one has been set. This would generally be used for conditioning a graph on known data without modifying the graph (see section 4.2.2.2).

Reading this property results in an error if no fixed value has been set. To determine if a fixed value has been set, use the hasFixedValue method (see section 5.2.3.3.1).

When setting this property on a single variable, the value must be a value included in the domain of the variable. The fixed value must be a value chosen from the domain of the variable. For example:

[23] If more than one domain element has identical marginal probabilities that are larger than for any other value, a single value from the domain is returned, chosen arbitrarily among these.

```
a = Discrete(1:10);
a.FixedValue = 3;
```

When setting this property on a variable array, the value must be an array of the same dimensions as the variable array, and each entry in the array must be an element of the domain.

Because the Input and FixedValue properties serve similar purposes, setting one of these overrides any previous use of the other. Setting the Input property removes any fixed value and setting the FixedValue property replaces the input with a delta function—the value 0 except in the position corresponding to the fixed value that had been set.

5.2.3.3 Methods

5.2.3.3.1 hasFixedValue

This method takes no arguments. When called for a single variable, it returns a boolean indicating whether or not a fixed-value is currently set for this variable. When called for a variable array, it returns a boolean array of dimensions equal to the size of the variable array, where each entry indicates whether a fixed value is set for the corresponding variable.

5.2.4 Bit

A Bit is a special kind of Discrete with domain [0 1].

5.2.4.1 Constructor

The Bit constructor can be used to create an N-dimensional collection of Dimple Discrete variables. Its constructor does not require a domain, since the domain is predetermined. The constructor takes only a variable-length list of matrix dimensions, where an empty list indicates a single Bit variable.

```
Bit([dimensions])
```

The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

5.2.4.2 Properties

5.2.4.2.1 Belief

Read-only. For a single Bit variable, the Belief property is a single number that represents the estimated marginal probability of the value one.

For an array of Bit variables, the Belief property is an array of numbers with size equal to the size of the variable array, with each value representing the estimated marginal probability of one for the corresponding variable.

5.2.4.2.2 Value

See section 5.2.3.2.2.

5.2.4.2.3 Input

Read-write. For setting the Input property on a single Bit variable, the value must be a single number in the range 0 to 1, which represents the normalized likelihood of the value one (see section 4.2.2.1). If $L(x)$ is the likelihood of the variable, the Input should be set to $\frac{L(x=1)}{L(x=0)+L(x=1)}$.

For setting the Input property on an array of Bit variables, the value must be an array of normalized likelihood values, where the array dimensions must match the dimensions of the array (or the portion of the array) being set.

5.2.4.2.4 FixedValue

See section 5.2.3.2.4.

5.2.4.3 Methods

5.2.4.3.1 hasFixedValue

See section 5.2.3.3.1.

5.2.5 Real

A Real variable represents a variable that takes values on the real line, or on a contiguous subset of the real line. The Real class corresponds to either a single Real variable or a multidimensional array of Real variables. All properties/methods can either be called for all elements in the collection or for individual elements of the collection.

5.2.5.1 Constructor

```
Real([domain], [dimensions])
```

All arguments are optional and can be used in any combination.

- domain specifies a bound on the domain of the variable. It can either be specified as a two-element array or a RealDomain object (see section 5.2.10). If specified as an array, the first element is the lower bound and the second element is the upper bound. -Inf and Inf are allowed values for the lower or upper bound, respectively. If no domain is specified, then a domain from $-\infty$ to ∞ is assumed.
- dimensions specify the array dimensions. The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

Examples:

- Real() specifies a scalar real variable with an unbounded domain.
- Real(4,1) specifies a 4x1 vector of real variables with unbounded domain.
- Real([-1 1]) specifies a scalar real variable with domain from -1 to 1.
- Real([-Inf 0], 4, 10, 2) specifies a 4x10x2 array of real variables, each with the domain from negative infinity to zero.
- Real(RealDomain(-pi, pi)) specifies a scalar real variable with domain from $-\pi$ to π.

5.2.5.2 Properties

5.2.5.2.1 Belief

Read-only. The behavior of this property for Real variables is solver specific. Some solvers do not support this property at all and will return an error when read. See section 5.6 for more detail on each of the supported solvers.

For the SumProduct solver, this property returns the estimated marginal distribution of the variable in the form of a NormalParameters object (see section 5.2.14), which includes a mean and precision value[24]. For an array of Real variables, this property returns a cell array of NormalParameters objects, each corresponding to the estimated marginal distribution of the corresponding variable. The results are undefined if called prior to running a solver.

5.2.5.2.2 Value

Read-only. The behavior of this property for Real variables is solver specific. Some solvers do not support this property at all and will return an error when read. See section 5.6 for more detail on each of the supported solvers.

For the SumProduct solver, the Value corresponds to the mean value of the belief.

5.2.5.2.3 Input

Read-write. For a Real variable, the Input property is expressed in the form of a FactorFunction object that can connect to exactly one Real variable. The list of available built-in FactorFunctions is given in section 4.1.4.6. Typically, an Input would use one of the standard distributions included in this list. In this case, it must be one in which all the parameters can be fixed to pre-defined constants. For the Gibbs and ParticleBP solvers, any such factor function may be used as an Input. For the SumProduct solver, however, only a Normal factor function may be used. Below is an example of setting the Input for a Real variable:

```
r = Real();
r.Input = FactorFunction('Normal', measuredMean, measurementPrecision);
```

Equivalently, a simpler alternative form for specifying the Input may be used. In this form the name of the factor followed by the arguments to the corresponding constructor are listed as elements of a cell array:

```
r = Real();
r.Input = {'Normal', measuredMean, measurementPrecision};
```

To remove an Input that had previously been set, the Input may be set to an empty array.

In the current version of Dimple, Inputs on Real variable arrays must be set one at a time, or all set to a single common value[25].

[24] For backward compatibility with Dimple version 0.04 and earlier, the Belief can be treated as a two-dimensional array where the first element is the mean and the second is the standard deviation.

[25] This restriction may be removed in a future version.

5.2.5.2.4 FixedValue

Read-write. The behavior of the FixedValue property for a Real variable is nearly identical to that of Discrete variables (see section 5.2.3.2.4). When setting the FixedValue of a Real variable, the value must be within the domain of the variable, that is greater than or equal to the lower bound and less than or equal to the upper bound. For example:

```
a = Real([-pi pi]);
a.FixedValue = 1.7;
```

Because the Input and FixedValue properties serve similar purposes, setting one of these overrides any previous use of the other. Setting the Input property removes any fixed value and setting the FixedValue property removes the input.

5.2.5.3 Methods

5.2.5.3.1 hasFixedValue

See section 5.2.3.3.1.

5.2.6 RealJoint

A RealJoint variable is a tightly coupled set of real variables that are treated by a solver as a single joint variable rather than a separate collection of variables. For example, in the SumProduct solver, the messages associated with RealJoint variables involve joint mean and covariance matrix rather than an individual mean and variance for each variable.

Like other variables, the RealJoint class can represent either a single RealJoint variable (representing a collection of real values) or an array of RealJoint variables.

5.2.6.1 Constructor

```
RealJoint(domain, [dimensions])
RealJoint(numElements, [dimensions])
```

The arguments are defined as follows:

- domain specifies a bound on the domain of the variable. It is specified by a RealJointDomain object (see section 5.2.11). If no domain is specified, then an unbounded domain is assumed and numElements must be specified instead.
- numElements specifies the number of joint real-valued elements. This argument is present only if the domain argument is not specified.
- dimensions specify the array dimensions (the array of individual RealJoint variables). The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

5.2.6.2 Properties

5.2.6.2.1 Belief

Read-only. The behavior of this property for RealJoint variables is solver specific. Some solvers do not support this property at all and will return an error when read. See section 5.6 for more detail on each of the supported solvers.

For the SumProduct solver, this property returns the estimated marginal distribution of the variable in the form of a MultivariateNormalParameters object (see section 5.2.15), which includes a mean vector and covariance matrix. For an array of RealJoint variables, this property returns a cell array of MultivariateNormalParameters objects, each corresponding to the estimated marginal distribution of the corresponding variable. The results are undefined if called prior to running a solver.

5.2.6.2.2 Value

Read-only. The behavior of this property for Real variables is solver specific. Some solvers do not support this property at all and will return an error when read. See section 5.6 for more detail on each of the supported solvers.

For the SumProduct solver, this property returns the mean vector, with dimension equal to the dimension of the RealJoint variable.

For an array of RealJoint variables, this property returns an array where the initial dimensions correspond to the dimensions of the variable array (or portion of the variable array), and the final dimension corresponds to the dimension of the RealJoint variable.

5.2.6.2.3 Input

Read-write. For a RealJoint variable, the Input property is expressed in one of two forms: either a single FactorFunction object that can connect to exactly one RealJoint variable, or a set of FactorFunction objects that can each connect to exactly one Real variable. The latter case corresponds to a likelihood function where each dimension is independent. The list of available built-in FactorFunctions is given in section 4.1.4.6. Typically, an Input would use one of the standard distributions included in this list. In this case, it must be one in which all the parameters can be fixed to pre-defined constants. For the Gibbs and ParticleBP solvers, any such factor function may be used as an Input. For the SumProduct solver, however, only a MultvariateNormal factor function or a set of Normal factor functions may be used.

Below is an example of setting the Input for a RealJoint variable using a single multivariate factor function:

```
r = Real();
r.Input = FactorFunction('MultivariateNormal', measuredMeanVector,
    measurementCovarianceMatrix);
```

Equivalently, a simpler alternative form for specifying the Input may be used. In this form the name of the factor followed by the arguments to the corresponding constructor are listed as elements of a cell array:

```
r = Real();
r.Input = {'MultivariateNormal', measuredMeanVector,
    measurementCovarianceMatrix};
```

To remove an Input that had previously been set, the Input may be set to an empty array.

To specify a set of univariate factor functions the value of this property must be a cell array of FactorFunction objects, one for each dimension of the RealJoint variable.

In the current version of Dimple, Inputs on RealJoint variable arrays must be set one at a time, or all set to a single common value[26].

5.2.6.2.4 FixedValue

Read-write. The behavior of the FixedValue property for a RealJoint variable is similar to that of Discrete variables (see section 5.2.3.2.4). When setting the FixedValue of a Real variable, the value must be within the domain of the variable. When setting a fixed value, the value must be in an array with dimension equal to the dimension of the RealVariable. For example:

```
a = RealJoint(4);
```

[26] This restriction may be removed in a future version.

```
a.FixedValue = [1.7, 2.0, 0, -1.2];
```

For an array of RealJoint variables, the fixed values of all variables may be set together. In this case, the initial dimensions of the input array must equal the dimensions of the variable array (or subset of the variable array) being set, while the final dimension must equal the dimension of the RealJoint variable.

Because the Input and FixedValue properties serve similar purposes, setting one of these overrides any previous use of the other. Setting the Input property removes any fixed value and setting the FixedValue property removes the input.

5.2.6.3 Methods

5.2.6.3.1 hasFixedValue

See section 5.2.3.3.1.

5.2.7 Complex

Complex is a special kind of RealJoint variable with exactly two joint elements.

The behavior of all properties and methods is identical to that of RealJoint variables, with the exception of a few methods (described below), which that refer directly to complex numerical values.

5.2.7.1 Constructor

```
Complex([domain], [dimensions])
```

The arguments are defined as follows:

- domain specifies the domain of the Complex variable using a ComplexDomain object (see 5.2.12). If no domain is specified, then an unbounded domain is assumed.
- dimensions specify the array dimensions (the array of individual Complex variables). The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

5.2.7.2 Properties

5.2.7.2.1 Belief

Read-only. See section 5.2.6.2.1.

5.2.7.2.2 Value

Read-only. This method behaves similarly to the Value property for a RealJoint variable, except the value returned is a single complex number. For an array of Complex variables, each entry in the returned array is a complex number.

5.2.7.2.3 Input

Read-write. See section 5.2.6.2.3.

5.2.7.2.4 FixedValue

Read-write. This method behaves similarly to the FixedValue property for a RealJoint variable, except that the fixed value is a single complex number. For an array of Complex variables, each entry in the array should be a complex number. For example:

a = Complex(); a.FixedValue = 5 + 1i*2;

5.2.7.3 Methods

5.2.7.3.1 hasFixedValue

See section 5.2.3.3.1.

5.2.8 FiniteFieldVariable

Dimple supports a special variable type called a FiniteFieldVariable, which represent finite fields with $N = 2^n$ elements. These fields find frequent use in error correcting codes. These variables are used along with certain custom factors that are implemented more efficiently for sum-product belief propagation than the alternative using discrete variables and factors implemented directly. See section 4.4 for more information on how these variables are used.

The behavior of all properties and methods is identical to that of Discrete variables.

5.2.8.1 Constructor

```
FiniteFieldVariable(primitivePolynomial, [dimensions])
```

The arguments are defined as follows:

- primitivePolynomial the primitive polynomial of the finite field. The format of the primitive polynomial follows the same definition used by MATLAB in the `gf` function. See the MATLAB help on the `gf` function for more detail.
- dimensions specify the array dimensions (the array of individual FiniteFieldVariable variables). The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

5.2.9 DiscreteDomain

The DiscreteDomain class represents a domain with a finite fixed set of elements. It is the type of Domain used by Discrete variables. DiscreteDomain objects are immutable.

5.2.9.1 Construction

```
DiscreteDomain(elementList)
```

The elementList argument is either a cell array or array of domain elements. Every entry of the array or cell array is considered an element of the domain, regardless of the number of dimensions it has. For a cell array, each object in the cell array is considered an element of the domain regardless of the object type. For a numeric array, every entry in the array must be numeric.

5.2.9.2 Properties

5.2.9.2.1 Elements

Read-only. This property returns the set of elements in the discrete domain in the form of a one-dimensional cell array.

5.2.10 RealDomain

The RealDomain class is used to refer to the domain of Real variables.

5.2.10.1 Constructor

```
RealDomain([lowerBound, [upperBound] ])
```

- lowerBound indicates the lower bound of the domain. The value must be a scalar numeric value. It may be set to -Inf to indicate that there is no lower bound. The default value is -Inf.
- upperBound indicates the upper bound of the domain. The value must be a scalar numeric value. It may be set to Inf to indicate that there is no upper bound. The default value is Inf.

5.2.10.2 Properties

5.2.10.2.1 LB

Read-only. This property returns the value of the lower bound. The default value is negative infinity.

5.2.10.2.2 UB

Read-only. This property returns the value of the upper bound. The default value is positive infinity.

5.2.11 RealJointDomain

The RealJointDomain class is used to refer to the domain of RealJoint variables.

5.2.11.1 Constructor

```
RealJointDomain(numDimensions);
RealJointDomain(listOfRealDomains);
RealJointDomain(numDimensions, realDomain);
```

- numDimensions indicates the number of dimensions in the domain of the RealJoint variable. If the form of constructor that specifies numDimensions is called, then all dimensions are assumed to be unbounded.
- listOfRealDomains is a comma-separated list or cell array of RealDomain objects, one for each dimension. Each RealDomain in the list indicates the domain for the corresponding dimension of the RealJoint variable.
- realDomain specifies a single RealDomain. If the number of dimensions is specified along with a single RealDomain, then that RealDomain is used for each dimension of the domain of the RealJoint variable.

5.2.11.2 Properties

5.2.11.2.1 NumElements

Read-only. Indicates the number of elements in the RealJointDomain, which corresponds to the number of dimensions of an associated RealJoint variable.

5.2.11.2.2 RealDomains

Read-only. Returns the collection of RealDomains that correspond to each dimension of the RealJointDomain.

5.2.12 ComplexDomain

The ComplexDomain class, a subclass of the RealJointDomain class, is used to refer to the domain of Complex variables.

5.2.12.1 Constructor

```
ComplexDomain();
ComplexDomain(listOfRealDomains);
ComplexDomain(realDomain);
```

- listOfRealDomains comma-separated list of exactly two RealDomain objects, one for each dimension. Each RealDomain in the list indicates the domain for the corresponding dimension of the Complex variable (real followed by imaginary). If no RealDomains are specified, then both dimensions are assumed to be unbounded.
- realDomain specifies a single RealDomain to apply to both dimensions of the domain of the Complex variable.

5.2.12.2 Properties

5.2.12.2.1 NumElements

Read-only. Indicates the number of elements in the ComplexDomain, which should always equal two.

5.2.12.2.2 RealDomains

Read-only. Returns the collection of RealDomains that correspond to each dimension of the ComplexDomain (real followed by imaginary).

5.2.13 FiniteFieldDomain

The FiniteFieldDomain class represents the domain of a FiniteFieldVariable. FiniteFieldDomain objects are immutable.

5.2.13.1 Construction

```
FiniteFieldDomain(primitivePolynomial)
```

- primitivePolynomial is an integer representation of the primitive polynomial of the finite field.

5.2.13.2 Properties

5.2.13.2.1 Elements

Read-only. This property returns the set of elements in the discrete domain in the form of a one-dimensional cell array.

5.2.13.2.2 PrimitivePolynomial

Read-only. This property returns the primitive polynomial associated with the domain using an integer representation.

5.2.13.2.3 N

Read-only. This property returns the number of bits in the finite field. The size of the Elements property is 2^N.

5.2.14 NormalParameters

The NormalParameters class is used to specify the parameters of a univariate Normal distribution, as used in the SumProduct solver.

5.2.14.1 Constructor

```
NormalParameters(mean, precision)
```

- mean is the mean value of the distribution.
- precision is the precision of the distribution, which is the inverse of the variance.

5.2.14.2 Properties

5.2.14.2.1 Mean

Read-only. Returns the mean value.

5.2.14.2.2 Precision

Read-only. Returns the precision value.

5.2.14.2.3 Variance

Read-only. Returns the variance (inverse of the precision).

5.2.14.2.4 StandardDeviation

Read-only. Returns the standard deviation (square root of the variance).

5.2.15 MultivariateNormalParameters

The MultivariateNormalParameters class is used to specify the parameters of a multivariate Normal distribution, as used in the SumProduct solver.

5.2.15.1 Constructor

```
MultivariateNormalParameters(meanVector, covarianceMatrix)
```

- meanVector indicates the mean value of each element in a joint set of variables. The value must be a one-dimensional numeric array.
- covarianceMatrix indicates the covariance matrix associated with the elements of a joint set of variables. The value must be a two-dimensional numeric array with each dimension identical to the length of the meanVector.

5.2.15.2 Properties

5.2.15.2.1 Mean

Read-only. Returns a vector of values, where each value indicates the mean value of the corresponding element in a joint set of variables.

5.2.15.2.2 Covariance

Read-only. Returns a two-dimensional array of values, representing the covariance matrix of a joint set of variables.

5.2.15.2.3 InformationVector

Read-only. Returns a vector of values representing the information matrix, defined as $\Sigma^{-1}\mu$, where Σ is the covariance matrix and μ is the mean vector.

5.2.15.2.4 InformationMatrix

Read-only. Returns a two-dimensional array of values representing the information matrix, defined as Σ^{-1}, where Σ is the covariance matrix.

5.3 Factors and Related Classes

5.3.1 Factor

The Factor class can represent either a single factor or a multidimensional array of factors. The Factor class is never created directly, but is the result of using the addFactor or addFactorVectorized (or related) methods on a FactorGraph.

5.3.1.1 Properties

5.3.1.1.1 Name

Read-write. When read, retrieves the current name of the factor or array of factors. When set, modifies the name of the factor to the corresponding value. The value set must be a string.

```
factor.Name = 'string';
```

When setting the Name, only one factor in an array may be set at a time. To set the names of an entire array of factors to distinct values, the setNames method may be used (see section 5.3.1.2.1).

5.3.1.1.2 Label

Read-write. All variables and factors in a Factor Graph must have unique names. However, sometimes its desirable to have variables or factors share similar strings when being plotted or printed. Users can set the Label property to set the name for display. If the Label is not set, the Name will be used for display. Once the label is set, the label will be used for display.

```
factor.Label = 'string';
```

```
factor.setLabel("string");
```

5.3.1.1.3 DirectedTo

Read-write. The DirectedTo property indicates a set of variables to which the factor is directed. The value may be either a single variable or a cell array of variables. The DirectedTo property is used by some solvers, and in some cases is required for proper operation of certain features. Such cases are identified elsewhere in this manual.

For example, if a factor F corresponds to a function $F(a, b, c, d)$, where a, b, c, and d are variables, then the factor is directed toward c and d if $\sum_{c,d} F(a, b, c, d)$ is constant for all values of a and b. In this case, we may set:

```
F.DirectedTo = {c, d};
```

In some cases, the set of DirectedTo variables can be automatically determined when a factor is created. In this case it need not be set manually by the user. This includes many built-in factors supported by Dimple. If this property is set by the user, then in the case of factors connected only to discrete variables, Dimple will check that the factor is in fact directed in the specified direction.

If the set of variables the factor are directed toward are part of a variable array, then these may be specified together in a single cell array. For example, if varArray is an array of variables, and a factor F is directed toward all of the variables in varArray, then we can set:

```
F.DirectedTo = varArray;
```

In the case of a vector of factors, we can identify the variables to which each factor is directed in a vectorized way. For example:

```
s = Discrete(domain,N);
fg.addFactorVectorized(factorFunction, s(1:(end-1)), s(2:end)).DirectedTo =
    s(2:end);
```

This example also shows that the DirectedTo property can be set directly on the result of the factor creation without assigning the factor to a named variable.

As a more complicated vectorized example, the following creates 12 factors, each of which contains 10 variables (5 from a and 5 from b). The first 2 of the 5 from a and the first from b are what the factor is directed to.

```
a = Bit(3,4,5);
b = Bit(3,4,5);
fg.addFactorVectorized(factorFunction, {a, [1 2]}, {b, [1 2]}).DirectedTo =
    {a(:,:,1:2), b(:,:,1)};
```

5.3.1.1.4 Score

Read-only. When read, computes and returns the score (energy) of the factor given a specified value for each of the neighboring variables to this factor. The score represents the energy of the factor given the specified variable configuration. The score value is relative, and may be arbitrarily normalized by an additive constant.

The value of each variable used when computing the Score is the Guess value for that variable (see section 5.2.2.1.5). If no Guess had yet been specified for a given variable, the value with the most likely belief (which corresponds to the Value property of the variable) is used[27].

The variable energy is normalized by the maximum input probability.

5.3.1.1.5 InternalEnergy

Read-only. (Only applies to the Sum Product Solver). When read returns:

$$InternalEnergy(a) = \sum_{\vec{x} \in \vec{X}} B_a(\vec{x}) * (-log(Weight(\vec{x})))$$

Where a is an instance of a Factor, X is the set of variables connected to a, Weight is the FactorTable entry for the specified set of variable values, and B_a is the belief of that factor node.

5.3.1.1.6 Bethe Entropy

Read-only. (Only applies to the Sum Product Solver). When read returns:

$$BetheEntropy(a) = - \sum_{\vec{x} \in domain(\vec{X})} B_a(\vec{x}) * log(B_a(\vec{x}))$$

Where a is an instance of a Factor, X is the set of variables connected to a, and B_a is the belief of that factor node.

```
double be = f.getBetheEntropy();
```

5.3.1.1.7 Belief

[27] For some solvers, beliefs are not supported for all variable types; in such cases there is no default value, so a Guess must be specified.

Read-only. (Only applies to the Sum Product Solver). To support the Bethe Free Energy property, Dimple provides getBelief associated with a Factor.

$$Belief_a(\vec{x}) = Weight(\vec{x}) \prod_{i=0}^{N} \mu_{X_i \to a}(x_i)$$

Where $\vec{x} \in domain(\vec{X})$ and $\vec{X}$ is the set of variables connected to the factor a.

```
b = f.Belief;
```

5.3.1.1.8 Ports

Read-only. Retrieves a cell array containing a list of Ports connecting the factor to its neighboring variables.

5.3.1.2 Methods

5.3.1.2.1 setNames

For an array of factors, the setNames method sets the name of each factor in the array to a distinct value derived from the supplied string argument. When called with:

```
factorArray.setName('baseName');
```

the resulting factor names are of the form: `baseName_vv0`, `baseName_vv1`, `baseName_vv2`, etc., where each factor's name is the concatenation of the base name with the suffix `_vv` followed by a unique number for each factor in the array.

5.3.1.2.2 invokeSolverSpecificMethod

```
factorArray.invokeSolverSpecificMethod('methodName', arguments);
```

For an array of factors, the invokeSolverSpecificMethod calls the specified solver-specific method on each of the factors in the array. Here, 'methodName' is a text string with the name of the solver-specific method, and arguments is an optional comma-separated list of arguments to that method. This method does not result in any return values. For solver specific methods that return results, use the invokeSolverSpecificMethodWithReturnValue method instead.

5.3.1.2.3 invokeSolverSpecificMethodWithReturnValue

```
returnArray = factorArray.invokeSolverSpecificMethodWithReturnValue('
    methodName', arguments);
```

For an array of factors, the invokeSolverSpecificMethodWithReturnValue calls the specified solver-specific method on each of the factors in the array, returning a return value for each variable. Here, 'methodName' is a text string with the name of the solver-specific method, and arguments is an optional comma-separated list of arguments to that method. The returnArray is a cell-array of the return values of the method, with dimensions equal to the dimensions of the factor array.

5.3.2 DiscreteFactor

When all variables connected to a Factor are discrete, a DiscreteFactor is created.

5.3.2.1 Properties

5.3.2.1.1 Belief

Read-only. The belief of a factor is the joint belief over all joint states of the variables connected to that factor. There are two properties that represent the belief in different ways: Belief and FullBelief. Reading the Belief property after the solver has been run[28] returns the belief in a compact one-dimensional vector that includes only values that correspond to non-zero entries in the factor table. This form is used because in some situation, the full representation over all possible variable values (as returned by the FullBelief property) would result in a data structure too large to be practical.

```
fb = myFactor.Belief;
```

The result is a vector of belief values, where the order of the vector corresponds to the order of the factor table entries. The order of factor table entries can be determined from the factor using:

[28] In the current version of Dimple, the Belief property is only supported for factors connected exclusively to discrete variables, and is supported only by the SumProduct solver. These restrictions may be removed in a future version.

```
ind = f.FactorTable.Indices
```

This returns a two-dimensional array, where each row corresponds to one entry in the factor table, and where each column-entry in a row indicates the index into the domain of the corresponding variable (where the order of the variable is the order used when the factor was created).

5.3.2.1.2 FullBelief

Read-only. Reading the FullBelief property after the solver has been run[29] returns the belief in a multi-dimensional array, where each dimension of the multi-dimensional array ranges over the domain of the corresponding variable (the order of the dimensions corresponds to the variable order used when the factor was created).

```
fb = myFactor.FullBelief;
```

5.3.3 FactorFunction

The FactorFunction class is used to specify a Dimple built-in factor function in a way that can be reused for creating multiple factors, and that allows specification of constructor arguments.

5.3.3.1 Constructor

```
FactorFunction(factorFunctionName, [constructorArguments])
```

- factorFunctionName is a string that indicates the name of the built-in factor function.
- constructorArguments is a variable-length comma-separated list of constructor arguments, whose interpretation is specific to the particular built-in factor function. If no arguments are needed, this list would be empty.

There are no available properties or methods in this class.

[29] In the current version of Dimple, the Belief property is only supported for factors connected exclusively to discrete variables, and is supported only by the SumProduct solver. These restrictions may be removed in a future version.

5.3.4 FactorTable

The FactorTable class is used to explicitly specify a factor table in lieu of Dimple creating one automatically from a factor function. This is sometimes useful in cases where the factor table is highly structured, but automatic creation would be time consuming due to a large number of possible states of the connected variables.

5.3.4.1 Constructor

```
FactorTable([ ( indexList, weightList ) | weightMatrix ], domains)
```

A FactorTable is constructed by specifying the table values in one of two forms, or by creating an all-zeros FactorTable to be filled in later using the set method. The first form, specifying an indexList and weightList, is useful for sparse factor tables in which many entires are zero and need not be included in the table. The second form, specifying a weightMatrix, is useful for dense factor tables in which most or all of the entries are non-zero.

- indexList is an array where each row represents a set of zero-based indices into the list of domain elements for each successive domain in the given set of domains.
- weightList is a one-dimensional array of real-valued entries in the factor table, where each entry corresponds to the indices given by the corresponding row of the indexList.
- weightMatrix is an N dimensional array of real-valued entries in the factor table. The number of dimensions, N, must correspond to the number of discrete domain elements given in the subsequent arguments, and the number of elements in each dimension must equal the number of elements in the corresponding domain element.
- domains is a comma-separated list of one or more discrete variable domains. Each domain either in the form of a DiscreteDomain object or a cell-array of the elements of the domain. From an existing variable, the domain can be obtained using the Domain property of that variable.

An example using the first method is:

```
a = Bit;
b = Bit;
c = Bit;
ft = FactorTable([0 0 0; 0 1 1; 1 0 1; 1 1 0], [1 .9 .5 .3], a.Domain, b.
    Domain, c.Domain);
```

An example using the second method is:

```
a = Discrete({'red','green','blue'});
b = Bit;
ft = FactorTable([.5 .4; .3 .1; 0 .4], a.Domain, b.Domain);
```

5.3.4.2 Properties

5.3.4.2.1 Indices

Read-write. When read, returns an array of indices of the factor table corresponding to entries in the factor table. Each row represents a set of zero-based indices into the list of domain elements for each successive domain in the given set of domains.

When written, replaces the previous array of indices with a new array. When writing using this property, the number of rows in the table must not change since this must equal the number of entires in the Weights. To change both Indices and Weights simultaneously, use the change method.

5.3.4.2.2 Weights

Read-write. When read, returns a one-dimensional array of real-valued entries in the factor table, where each entry corresponds to the indices given by the corresponding row of the indexList.

When written, replaces the previous array of weights. When writing using this property, the number of entries must not change since this must equal the number of rows in the Indices. To change both Indices and Weights simultaneously, use the change method.

5.3.4.2.3 Domains

Read-only. Returns a cell-array of DiscreteDomain objects, each of which represents the corresponding domain specified in the constructor, in the same order as specified in the constructor.

5.3.4.3 Methods

5.3.4.3.1 set

This method allows setting individual entries in the factor table. For each entry to be set, the domain values (not indices) are specified followed by the weight value. To set a single entry, the domain values are specified in a comma-separated list, followed by the weight value. To set multiple entries in a single call, a cell array of such comma-separated lists are specified.

An example of setting a single entry:

```
ft.set(’red’, 1, 0.45);
```

An example of setting multiple entries:

```
ft.set({’red’, 1, 0.45}, {’blue’, 0, 0.75});
```

5.3.4.3.2 get

This method retrieves the weight associated with a particular entry in the factor table. The entry is specified by a comma-separated list of domain values (not indices). For example:

```
w = ft.get(’red’, 1);
```

5.3.4.3.3 change

This method allows simultaneously replacing both the array of Indices and Weights in the factor table.

```
ft.change(indexList, weightList);
```

The arguments indexList and weightList are exactly as described in the FactorTable constructor.

5.4 Options

Dimple provides an option mechanism used to configure the runtime behavior of various aspects of the system. This section describes the Dimple option system and how it is used. Individual options are described in more detail later in this document.

Options were introduced in version 0.07 and replace earlier mechanisms based on solver-specific method calls. Those methods are now deprecated and will be removed in a future release. Users with existing Dimple code using such methods should switch to using options as soon as it is convenient to do so.

5.4.1 Option Keys

Options are key/value pairs that can be set on Dimple factor graph, variable, or factor objects to configure their behavior. An option key uniquely identifies an option, along with its type and default value.

In the MATLAB API, option keys are represented by unique strings of the form '*OptionClass.optionName*' (a complete list of supported option keys is returned by the dimpleOptions() function). For instance:

```
graph.setOption('BPOptions.iterations', 12);
```

Specifying a string that does not correspond to a known option key will result in a runtime error when key is used to set or look up an option value.[30]

5.4.2 Setting Options

Options may be set on any FactorGraph, Factor, or Variable object or their solver-specific counterparts. Options may also be set on the DimpleEnvironment object, which is described in more detail below. Options set on graphs will be applied to all factors, variables, and subgraphs contained in the graph unless overridden on one of those members. Likewise options set on a model object will be applied to an associated Solver object to it unless overridden directly in the Solver object. In most cases, it should not be necessary to set options directly on Solver objects.

Options can be set using the setOption method of the Node class. For example:

```
graph.setOption('BPOptions.damping', .9);
```

When applied to an array object the option will be set on each. For example:

[30] Internally option keys are represented using singleton Java objects, which are looked up by the string keys. If you have a reference to the Java IOptionKey object, you can use it in place of the string throughout this interface. Consult the Java version of this manual for further details.

```
vars = Real(2,2);
vars.setOption('BPOptions.damping', .9);
```

In this case, to apply a distinct value to each element of the array, the values are specified in a cell-array of the same dimensions as the array object:

```
vars = Real(2,2);
vars.setOption('BPOptions.damping', {.7 .8; .6 .9});
```

Multiple options may be set at the same time using the setOptions method. This takes arguments in one of the following forms:

- A comma-separated list containing alternating option keys and values.

```
nodes.setOptions('BPOptions.iterations', 10, ...
                  'BPOptions.damping' , .9);
```

- A vector cell array containing alternating option keys and values.

```
nodes.setOptions({'BPOptions.iterations', 10, ...
                  'BPOptions.damping' , .9});
```

- A nx2 cell array where each row contains a key and value.

```
nodes.setOptions({'BPOptions.iterations', 10; ...
                  'BPOptions.damping' , .9});
```

- A cell array with dimensions matching the dimensions of the left hand side where each cell contains one of the above two forms.

```
options = cell(2,2);
options{1,1} = {'BPOptions.iterations', 10; ...
                'BPOptions.damping', .85};
options{2,2} = {'BPOptions.iterations', 12};
nodes.setOptions(options);
```

All of these methods will ensure that the type of the option values are appropriate for that key and may also validate the value. For instance when setting the BPOptions.damping option, the value must be a double in the range from 0.0 to 1.0. If a value is not valid for its key an OptionValidationException will be thrown.

Options may be unset on any object on which they were previously set using the unset method:

```
graph.unsetOption('BPOptions.damping');
```

All options may be unset on an object using the clearLocalOptions method:

```
graph.clearLocalOptions();
```

5.4.3 Looking up Option Values

There are a number of methods for retrieving option values from objects on which they can be set. Most users will only need to use these to debug their option settings.

The option value that applies to a given object is determined hierarchically, based on an order that depends on the structure of the graph. An option value specified at any level applies to all objects below it in the hierarchy, unless specifically specified for an object lower in the hierarchy. At any level, the option value overrides the value specified at a higher level. When querying an object to determine what option value will be used, the hierarchy is searched in the following order[31]:

1. Search the object itself.
2. If the object is a solver object, next look at the corresponding model object.
3. If the object has a parent graph, then recursively search that graph, otherwise the DimpleEnvironment for that object will be searched (there is usually only one environment).

There are two methods for looking up option values:

- getOption(key) - returns the values of the specified option as determined by the above lookup rules or else the option's default value if not set. If invoked on an object array then this will return a cell array containing the values. For example:

```
>>> graph.getOption('BPOptions.damping')
ans =
    0.9000

>>> vars.getOption('BPOptions.damping')
ans =
    [0.7000]    [0.6000]
    [0.8000]    [0.9000]
```

[31] The algorithm is actually slightly more complicated than this but the details should only matter to those implementing custom factors or solvers. For details see the documentation for EventSourceIterator in the HTML Java API documentation.

- getLocalOptions() - returns a cell array containing keys and values of options that are set directly on that object in the same format accepted by the setOptions method described in the previous section. For example:

```
>>> graph.getLocalOptions()
ans =
    'BPOptions.damping'     [0.9000]
    'BPOptions.iterations'      10
```

5.4.4 Option Initialization

While option values are visible as soon as they are set on an object, they may not take effect until later because internal objects that are affected by the change may have cached state based on the previous settings, or may not yet exist. The documentation for individual options should indicate when changes to the settings are incorporated, but in most cases that will happen when the initialize() method is called on the affected object. Since this happens automatically when invoking the FactorGraph.solve() method, users will often not have to be concerned with this detail. But if you performing other operations, such as directly calling FactorGraph.iterate(), then you will probably need to invoke FactorGraph.initialize() for modified option settings to take effect.

5.4.5 Setting Defaults on the Dimple Environment

Sometimes you may want to apply the same default option settings across multiple graphs. While you can simply set the options on all of the graphs individually, another choice is to set it on the DimpleEnvironment object. The DimpleEnvironment holds shared state for a Dimple session. Typically there will be only one instance of this class. Because the environment is the last place searched for option lookup, you can use it as a place to set default values of options to override those defined by the option keys.

You can obtain a reference to the active global environment using the static DimpleEnvironment.active() method: and set default option values on it. For instance, to globally enable multithreading for all graphs, you could write:

```
env = DimpleEnvironment.active();
env.setOption('SolverOptions.enableMultithreading', true);
```

5.5 Schedulers

A scheduler defines a rule that determines the update schedule of a factor graph when performing inference. This section describes all of the built-in schedulers available in Dimple.

Each scheduler is applicable only to a certain subset of solvers. For the BP solvers (other than the Junction Tree solvers, that is, SumProduct, MinSum, and ParticleBP), the following schedulers are available:

Name	Description
DefaultScheduler	Same as the TreeOrFloodingScheduler, which is the default if no scheduler or custom schedule is specified.
TreeOrFloodingScheduler	The solver will use either a Tree Schedule or a Flooding Schedule depending on whether the factor-graph contains cycles. In a nested graph, this choice is applied independently in each subgraph. If the factor-graph is a tree, the scheduler will automatically detect this and use a Tree Schedule. In this schedule, each node is updated in an order that will result in the correct beliefs being computed after just one iteration. If the entire graph is a tree, NumIterations should be set to 1, which is its default value. If the factor-graph is loopy, the solver will instead use a Flooding Schedule (as described below).
TreeOrSequentialScheduler	The solver will use either a Tree Schedule (as described above) or a Sequential Schedule (as described below) depending on whether the factor-graph contains cycles. In a nested graph, this choice is applied independently in each subgraph.
FloodingScheduler	The solver will apply a Flooding Schedule. For each iteration, all variable nodes are updated, followed by all factor nodes. Because the graph is bipartite (factor nodes only connect to variable nodes, and vice versa), the order of update within each node type does not affect the result.
SequentialScheduler	The solver will apply a Sequential Schedule. For each factor node in the graph, first, for each variable connected to that factor, the edge connecting the variable to the factor is updated; then the factor node is updated. The specific order of factors chosen is arbitrary, and depends on the order that factors were added to the graph.
RandomWithoutReplacementScheduler	The solver will apply a Sequential Schedule with the order of factors chosen randomly without replacement. On each subsequent iteration, a new random order is chosen. Since the factor order is chosen randomly with replacement, on each iteration, each factor will be updated exactly once.

RandomWithReplacementScheduler	The solver will apply a Sequential Schedule with the order of factors chosen randomly with replacement. On each subsequent iteration, a new random order is chosen. The number of factors updated per iteration is equal to the total number of factors in the graph. However, since the factors are chosen randomly with replacement, not all factors are necessarily updated in a single iteration, and some may be updated more than once.

For the JunctionTree and JunctionTreeMAP solvers, only a Tree Schedule will be used. When using these solvers, the Scheduler setting will be ignored.

In a nested graph, for most of the schedulers listed above (except for the random schedulers), the schedule is applied hierarchically. In particular, a subgraph is treated as a factor in the nesting level that it appears. When that subgraph is updated, the schedule for the corresponding subgraph is run in its entirety, updating all factors and variables contained within according to its specified schedule.

It is possible for subgraphs to be designated to use a schedule different from that of its parent graph. This can be done by specifying either a scheduler or a custom schedule for the subgraph prior to adding it to the parent graph. For example:

```
SubGraph.Scheduler = 'SequentialScheduler';
ParentGraph.addFactor(SubGraph, boundaryVariables);
ParentGraph.Scheduler = 'FloodingScheduler';
```

For the TreeOrFloodingScheduler and the TreeOrSequentialScheduler, the choice of schedule is done independently in the outer graph and in each subgraph. In case that a subgraph is a tree, the tree scheduler will be applied when updating that subgraph even if the parent graph is loopy. This structure can improve the performance of belief propagation by ensuring that the effect of variables at the boundary of the subgraph fully propagates to all other variables in the subgraph on each iteration.

For the RandomWithoutReplacementScheduler and RandomWithReplacementScheduler, if these are applied to a graph or subgraph, the hierarchy of any lower nesting layers is ignored. That is, the subgraphs below are essentially flattened prior to schedule creation, and any schedulers or custom schedules specified in lower layers of the hierarchy are ignored.

Because of the differences in operation between the Gibbs solver and the BP based solvers, the Gibbs solver supports a distinct set of schedulers. For the Gibbs solver, the following schedulers are available:

Name	Description
GibbsDefaultScheduler	Same as the GibbsSequentialScanScheduler, which is the default when using the Gibbs solver.
GibbsSequentialScanScheduler	The solver will apply a Sequential Scan Schedule. For each scan, each variable is resampled in a fixed order. The specific order of variables chosen is arbitrary, and depends on the order that variables were added to the graph.
GibbsRandomScanScheduler	The solver will apply a Random Scan Schedule. Each successive variable to be resampled is chosen randomly with replacement. The number of variables resampled per scan is equal to the total number of variables in the graph, but not all variables are necessarily resampled in a given scan, and some may be resampled more than once.

Because of the nature of the Gibbs solver, the nested structure of a graph is ignored in creating the schedule. That is, the graph hierarchy is essentially flattened prior to schedule creation, and only the scheduler specified on the outermost graph is applied.

Schedulers are not applicable in the case of the LP solver.

5.6 Solvers

5.6.1 Solver-Specific Options

Each solver supports a number of options specific to that solver. These are described in the following sections. Solver-specific options may be used to configure how overall inference works for that solver or may be used to configure the behavior of individual factors or variables for that solver. Solver-specific options are typically set on the applicable model object (factor graph, factor, or variable) and the values will be observed and used to configure the corresponding solver objects:

```
model-object.setOption('SolverOptionClass.optionName', option-value)
```

Solver-specific options may be set at any time, even before the solver for a graph has been specified. Options that are not applicable to the object on which it is set or that are not applicable to the active solver will simply be ignored. For more details on the options mechanism see subsection 5.4 of this document.

5.6.2 Solver-Specific Methods

Each solver also may support solver-specific methods, which are described in the following sections. As with options, solver-specific methods may be available for various objects: a factor-graph, variable, or factor. In each case, to call a solver-specific method, the method is applied to the solver object, returned by the Solver property. For example:

```
factorGraph.Solver.solverSpecificMethod(arguments);
```

```
variable.Solver.solverSpecificMethod(arguments);
```

```
factor.Solver.solverSpecificMethod(arguments);
```

Some solver-specific methods return results, while others do not. Some solver-specific methods require arguments, while others do not. If no arguments are needed, the parentheses are optional.

In some cases it is convenient to call solver-specific methods on each element in an array of objects. Utility methods are provided for this purpose. Specifically, to call a solver-specific method that has no return value:

```
objectArray.invokeSolverSpecificMethod('methodName', arguments);
```

In this case, 'methodName' is a text string with the name of the solver-specific method, and arguments is an optional comma-separated list of arguments to that method.

To call a solver specific method that has a return value:

```
returnArray = objectArray.invokeSolverSpecificMethodWithReturnValue('
    methodName', arguments);
```

In this case, returnArray is a cell-array of the return values of the method, with dimensions equal to the dimensions of the object array.

5.6.3 Common Options

A few options are applicable to multiple solvers and are therefore described in this subsection.

5.6.3.1 SolverOptions.enableMultithreading

Type boolean
Default false
Affects graph
Description Controls whether to use multithreading for this solver. Multithreading is currently only supported by the MinSum and SumProduct solvers but will eventually be implemented in others. This value will be ignored if not applicable.

5.6.3.2 DimpleOptions.randomSeed

Type 64-bit integer
Default N/A
Affects graph
Description When set, this option specifies a random seed that may be used by solvers that use a random number generator. The seed will only be used if explicitly set; the default value is not used. This can be used to ensure repeatable behavior during testing or profiling but should not be used for normal operation.

5.6.4 Common Methods

There are also some methods that are common to all solvers. These are:

5.6.4.1 getMultithreadingManager

Dimple users can retrieve a MultithreadingManager on which to perform additional actions.

```
fg.Solver.getMultithreadingManager()
```

Users can configure both the multithreading mode and the number of workers using the MultithreadingManager.

5.6.4.1.1 Multithreading Modes Dimple provides various multithreading algorithms that have different speed advantages depending on the size of the user's graph and FactorTables. In the future Dimple should be modified to automatically detect the best threading algorithm. Currently, however, it defaults to the ”Phase” multithreading mode and requires the user manually set the mode to change this. For a given graph, users can try both modes and see which is faster.

The currently supported multithreading modes are:

- Phase - Divides the schedule into ”phases” where each phase contains schedule entries that are entirely independent of one another. These phases are then easy to parallelize.
- SingleQueue - Uses a single queue and a dependency graph to pull off work for each thread on the fly.

The following methods can be used for getting and setting modes:

- fg.Solver.getMultithreadingManager().getModes() - Returns a Java array of enums specifying the valid modes.
- fg.Solver.getMultithreadingManager().setMode(ModeName) - Allows users to set the mode by string. Currently ”Phase” or ”SingleQueue” will work.
- fg.Solver.getMultithreadingManager().setMode(enum) - Allows users to set the mode by the enums returned by the getModes method.

5.6.4.1.2 Setting Number of Threads and Workers Dimple provides a ThreadingPool as a singleton for multithreading. It sets the number of threads in this pool to the number of virtual cores in the user's machine by default. Users can override this default value. In addition, Dimple allows users to specify the number of ”workers” for a given FactorGraph. This ”NumWorkers” is also set to the number of virtual cores on the user's machine by default. Whereas NumThreads specifies how many threads are in the threadPool, NumWorkers specifies how work is divided up across the graph. These workers are run by the thread pool. Best performance is achieved when NumWorkers and NumThreads are the same. However, NumThreads is global and shared by all graphs where NumWorkers is specific to a given FactorGraph.

The following methods can be used to change number of workers:

- fg.Solver.getMultithreadingManager().getNumWorkers()
- fg.Solver.getMultithreadingManager().setNumWorkers(num)
- fg.Solver.getMultithreadingManager().setNumWorkersToDefault()

The following global methods can be used to set the number of threads in the ThreadPool

- getDimpleNumThreads()
- setDimpleNumThreads(numThreads)
- setDimpleNumThreadsToDefault()

5.6.5 Common Belief Propagation Options

There are a number of options that are applicable to multiple solvers that are based on some form of message-passing belief propagation. These include the Sum-Product, Min-Sum, Particle BP, and Junction Tree solvers. These options are defined in the BPOptions class. The following options are supported:

5.6.5.1 BPOptions.iterations

Type integer
Default 1
Affects graph
Description Controls how many iterations to perform when running solve(). This is not applicable to all solvers. It is currently only used by the SumProduct, MinSum and ParticleBP solvers. It only makes sense to set this to a value greater than one if the graph is not singly connected or ”loopy”, that is when there is more than one unique path between two or more nodes in the graph. You can tell if a graph is loopy using the FactorGraph method isForest(), which will be false if the graph is not singly connected.

5.6.5.2 BPOptions.damping

Type double
Default 0.0
Affects variables and factors
Description The belief propogation based solvers supports damping, in which messages are damped by replacing each message by a weighted sum of the computed message value and the previous value of that message (when the corresponding edge was most-recently updated). In the current version of Dimple, damping is supported only in discrete variables and factors that connect only to discrete variables[32].

The damping parameter specifies a weighting value in the range 0 through 1:

$$\text{message} = \text{computedMessage} \cdot (1 - D) + \text{previousMessage} \cdot D$$

where D is the damping value. So that a value of 0 means that the previous message will not be considered, effectively turning off damping.

This option applies the same damping parameter to all edges connected to the variable or factor on which it is set. If you want different values for different edges, you need to use the BPOptions.nodeSpecificDamping option.

5.6.5.3 BPOptions.nodeSpecificDamping

Type	double vector
Default	*empty*
Affects	variables and factors
Description	This is the similar to the BPOptions.damping option but allows you specify different weights for different edges. Unlike the simple damping option, this usually makes no sense to set on the graph itself since factors and variables will typically have different numbers and arrangements of edges. The value must either be an empty list, indicating that damping should be turned off, or a list of weights with the same length as the number of siblings of the affected variable and factor. The damping weights will be applied in the order in which the siblings are declared. This option takes precedence over the simple damping option if both are specified for the same node.

5.6.5.4 BPOptions.maxMessageSize

Type	integer
Default	integer max
Affects	discrete factors
Description	This specifies the maximum size of the outgoing messages on the discrete factors on which it is set. If this number K is less than the full size of the message (i.e. the size of the domain of the target variable), then only the K-best values – those with the largest weights – will be included in the message. This can results in a faster but more approximate form of inference and is most suited to graphs with very large-dimension variables. IMPORTANT: k-best and damping are not compatible with each other[33]

5.6.5.5 BPOptions.updateApproach

Type string

Default AUTOMATIC

Affects discrete factors

Description This option controls which update algorithm is applied to discrete factors. The option can take one of three values:

- NORMAL - Perform updates using just the factor tables. Do not use the optimized update technique.
- OPTIMIZED - Use the optimized update algorithm. Note that factors that have only one edge, or factors that do not have all of their edges updated simultaneously by the schedule, ignore this setting and use the normal approach.
- AUTOMATIC - Automatically determine whether to use the optimized algorithm. The automatic selection algorithm can be tuned through the BPOptions.automaticExecutionTimeScalingFactor and BPOptions.automaticMemoryAllocationScalingFactor options.

5.6.5.6 BPOptions.automaticExecutionTimeScalingFactor

Type double

Default 1.0

Affects discrete factors

Description This option is an execution time scaling factor used when the BPOptions.updateApproach option is set to AUTOMATIC. It controls how execution time costs are weighed. The value must be a positive number.

5.6.5.7 BPOptions.automaticMemoryAllocationScalingFactor

Type double

Default 10.0

Affects discrete factors

Description This option is an memory allocation scaling factor used when the BPOptions.updateApproach option is set to AUTOMATIC. It controls how memory allocation costs are weighed. The value must be a positive number.

5.6.5.8 BPOptions.optimizedUpdateSparseThreshold

Type double

Default 1.0

Affects discrete factors

Description This option controls the representation of the auxiliary tables used by the optimized update algorithm, which is controlled through the BPOptions.updateApproach. Internally, the optimized algorithm creates multiple factor tables to perform the update. This option specifies a density, below which an auxiliary table uses a sparse representation. It must be a number in the range [0.0, 1.0]. The value 1.0 (the default), indicates that a sparse representation should be used if there are any zero-entries in the table. The value 0.0 will prevent the sparse representation from being used entirely. Sparse tables typically decrease execution time, but they use more memory. When the update approach is set to AUTOMATIC, this option impacts both the execution time and memory allocation estimates used to choose the update approach.

5.6.6 Sum-Product Solver

Use of the sum-product solver is specified by calling:

```
fg.Solver = 'SumProduct';
```

If no solver is specified, the SumProduct solver is used by default.

The SumProduct solver supports both discrete and continuous variables. The SumProduct solver uses the sum-product form of belief propagation to perform inference on a graph. For discrete variables, each message to or from a factor is in the form of a vector of length equal to the domain size of the variable. For continuous variables, messages are represented using a Gaussian parameterization. In some cases, this is an approximation to the exact message. For Real variables, a message is in the form of a pair of values representing the mean and variance of the corresponding Normal distribution. For Complex and RealJoint variables, a message is in the form of a vector and matrix, representing the mean and covariance of the corresponding multivariate Normal distribution.

While the Gaussian representation of messages for continuous variables is sometimes an approximation, there are some specific built-in factors for which exact Gaussian messages are computed. This can be done when a factor preserves the Gaussian form of the distribution on each edge. The following table is a list of such lists built-in factors. See section 5.9 for more information on built-in factors.

Built-in Factor	Variable Types	Notes
Normal	Real	Applies only if mean and precision parameters are constants and all connected variables are Real and unbounded[34].
MultivariateNormal	Complex or RealJoint	All connected variables must be RealJoint or Complex and unbounded.
Sum	Real	All connected variables must be Real and unbounded.
Subtract	Real	All connected variables must be Real and unbounded.
Negate	Real	All connected variables must be Real and unbounded.
ComplexSum	Complex	All connected variables must be Complex and unbounded.
ComplexSubtract	Complex	All connected variables must be Complex and unbounded.
ComplexNegate	Complex	All connected variables must be Complex and unbounded.
RealJointSum	RealJoint	All connected variables must be RealJoint of the same dimension and unbounded.
RealJointSubtract	RealJoint	All connected variables must be RealJoint of the same dimension and unbounded.

[34] Unbounded means that the domain of the variable must not have finite upper or lower bounds

Built-in Factor	Variable Types	Notes
RealJointNegate	RealJoint	All connected variables must be RealJoint of the same dimension and unbounded.
Product	Real, Constant	Applies only if the product is one unbounded Real variable times one scalar constant to produce an unbounded Real variable.
MatrixRealJoint VectorProduct	RealJoint, Constant	Applies only if the product is one unbounded RealJoint variable times a constant matrix to produce an unbounded RealJoint variable.
LinearEquation	Real	Linear equation. All connected variables must be Real and unbounded.

For factors that are neither discrete-only or listed in the above table, an approximate computation is used in computing messages from such a factor. This includes any factor that connects to both discrete and continuous variables as well as factors that connect only to continuous variables but do not appear in the list above. The approximate method is sample based and uses Gibbs sampling to sample from the factor, allowing approximate messages to be computed from the sample statistics. Several methods described below allow control over the behavior of these sampled factors.

For discrete-only factors, two factor update algorithms are available: normal and optimized. The optimized algorithm can be applied only to factors with more than one edge, and only when the schedule updates all of the factor's edges simultaneously. The optimized algorithm computes the outbound message update with fewer operations than the normal algorithm, which can decrease execution time; however, it also uses more memory and increases initialization time. Several options, described below, influence which algorithm is used. Key among them is the updateApproach option, which can be set to normal, optimized, or automatic. When set to automatic, Dimple makes an estimate of the memory usage and execution time of each algorithm in order to select one.

5.6.6.1 GibbsOptions for Sampled Factors

Factors connected to continuous variables that do not support exact message computation, instead use a sampled approximation (see section 5.6.6) where the sampling is performed using the Gibbs solver.

For all such factors in a graph, you may set any of the Gibbs solver options described in paragraph 5.6.9.1 to control how the sampling will be done. The most important of these options have different default values when used with Sum-Product. This is accomplished by setting the options on the solver graph object when it is constructed. In order to override these defaults, it is necessary to set them on the solver graph (to apply to all such factors, using `graph.Solver.setOption(...)`), or on the factor specific factor object (to apply to a single factor, using `factor.setOption(...)`).

These options and their SumProduct-specific default values are:

- GibbsOptions.numSamples: 1000
- GibbsOptions.burnInScans: 10
- GibbsOptions.scansPerSample: 1

5.6.7 Min-Sum Solver

Use of the MinSum solver is specified by calling:

```
fg.Solver = 'MinSum';
```

Unlike the Sum-Product solver, the Min-Sum solver supports only discrete variables. It only uses the standard BP Options described in Common Belief Propagation Options.

5.6.8 Junction Tree Solver

There are two distinct forms of Junction Tree solver in Dimple: the sum-product form used for computing exact marginal variable beliefs and the min-sum form used for computing MAP values. The Junction Tree solvers support only discrete variables.

Use of one of the two forms of Junction Tree solver by calling either:

```
fg.Solver = 'JunctionTree';
```

for the sum-product version or

```
fg.Solver = 'JunctionTreeMAP';
```

for the min-sum version.

The Junction Tree solvers are useful for performing exact inference on loopy discrete graphical models for which the standard sum-product or min-sum algorithms will only produce approximate results. This works by transforming the model into a corresponding non-loopy model, building a proxy solver layer that connects the original model to the transformed version and doing inference on that model. If your model already is non-loopy then you can simply use sum-product or min-sum directly for exact inference. To test to see if a graph is loopy or not, use isForest():

```
useJunctionTree = ~fg.isForest();
```

One significant limitation when using this solver is that the cost of inference and amount of memory needed to store the factor tables is proportional to the size of the tables, which in turn is exponential in the number of variables represented in a table. The Junction Tree algorithm may be unable to determine an equivalent tree structure that has small enough factors either to fit in memory or to perform inference on in an acceptable amount of time. The typical failure mode in such cases is to get an OutOfMemoryError when attempting to run solve. The Junction Tree algorithm works best when used with smaller graphs or larger graphs that have few loops or are loopy but long and narrow.

5.6.8.1 Junction Tree Options

The following options are available for use with both versions of the Junction Tree solver.

Because exact inference can be done in a single iteration, the Junction Tree solver fixes the number of iterations to one and will ignore attempts to set it to another value. Because damping would result in inexact inference, the Junction Tree solver does not provide options for using damping.

5.6.8.1.1 JunctionTreeOptions.useConditioning

Type boolean

Default false

Affects graph

Description Specifies whether to use conditioning when constructing the transformation of the model. When true, then any variables in the model that have a fixed value will be disconnected from the rest of the graph in the transformed version and its value will be incorporated in the factors of the transformed model. This will produce a more efficient transformed model when there are fixed values in the original model. Using this will require that a new transformation be computed every time a fixed value changes.

5.6.8.1.2 JunctionTreeOptions.maxTransformationAttempts

Type integer

Default 1

Affects graph

Description Specifies the maximum number of times the junction tree transformer should try to determine an optimal transformation. Each attempt uses a greedy ”variable elimination” algorithm using a randomly chosen cost function and random choices to break ties, so more iterations could produce a more efficient tree transformation.

5.6.9 Gibbs Solver

Use of the Gibbs solver is specified by calling:

```
fg.Solver = 'Gibbs';
```

The Gibbs solver supports both discrete and continuous variables.

This solver performs Gibbs sampling on a factor graph. It supports a variety of output information on the sampled graph, including the best joint sample (lowest energy), marginals of each variable (discrete variables only), and a full set of samples for a user-selected set of variables. The solver supports both sequential and randomized scan, and it supports tempering with an exponential annealing schedule.

The Gibbs solver supports several user selectable generic samplers (those that don't require specific conjugacy relationships). The following table lists the available generic samplers, and the variable types supported by each.

Sampler	Variable Type	Description
CDFSampler	Discrete[35]	Samples from the full conditional distribution of the variable. This is the default sampler for discrete variables.
SliceSampler	Real[36]	Slice sampler using the doubling procedure. See Neal, Slice Sampling (2000). This is the default sampler for real variables.
MHSampler	Discrete & Real	Metropolis-Hastings sampler. For discrete variables, the default proposal kernel is uniform over values other than the current value. For real variables, the default proposal kernel is Normal with standard deviation 1 (the standard deviation is user settable). Alternate proposal kernels are also available (see below).
SuwaTodoSampler	Discrete	Suwa-Todo sampler. See Suwa, Todo, Markov Chain Monte Carlo Method without Detailed Balance (2010).
BlockMHSampler	Discrete & Real	Block Metropolis-Hastings sampler. Allows block proposals for a collection of more than one variable at a time. In the current version of Dimple, there are no built-in general purpose block proposal kernels. To use this sampler, a custom block proposal kernel must be written in Java, and used by specifying a block schedule entry that references this proposal kernel. See section 5.6.9.8.1 for more information.

[35] In this table, Discrete support implies any of the discrete variable types, including Discrete and Bit.

[36] In this table, Real support implies any of the continuous variable data types, including Real, Complex, and RealJoint.

In cases where the factors of the graph support a conjugate distribution, the solver will automatically determine this and use the appropriate conjugate sampler. The following table lists the supported conjugate samplers and the corresponding factors that support them. The corresponding sampler will be used for a given variable if all of the edges connected to that variable support the same sampler[37].

Sampler	Built-in Factor	Edge
BetaSampler	Beta	value
	Binomial	ρ
DirichletSampler	Dirichlet	value
	Categorical	α
	DiscreteTransition	α
GammaSampler	Gamma	value, β
	NegativeExpGamma	β
	Normal	τ
	LogNormal	τ
	Poisson	λ
	CategoricalUnnormalizedParameters	α
	DiscreteTransitionUnnormalizedParameters	α
NegativeExpGammaSampler	NegativeExpGamma	value
	CategoricalEnergyParameters	α
	DiscreteTransitionEnergyParameters	α
NormalSampler	Normal	value, μ
	LogNormal	μ

Additionally, conjugate sampling is supported across a subset of applicable deterministic functions. In the current version of Dimple, this includes the following factors:

Factor Function	Edges	Condition
Multiplexer	in, out	If any of the *in* variables would support the same conjugate sampler as the *out* variable, then those *in* variables will use conjugate sampling.

The Gibbs solver automatically performs block-Gibbs updates for variables that are deterministically related. The Gibbs solver automatically detects deterministic relationships associated with built-in deterministic factor functions (see section 5.9 for a list of these functions).

For user-defined factors specified by MATLAB factor functions or factor tables, the Gibbs solver will detect if they are deterministic functions as along as the factor is marked as the directed outputs are indicated using the DirectedTo property, as described in section 5.3.1.1.3.

The Gibbs solver also automatically performs block-Gibbs updates for certain built-in factors that Gibbs sampling on individual variables would fail due to the dependencies between variables imposed by the factor. In these cases, a custom proposal distribution ensures that

[37] Additionally, for the conjugate sampler to be used, the domain of the variable must not be bounded to a range smaller than the natural range of the corresponding distribution.

proposals are consistent with the constraints of the factor. However, the custom proposals are not assured to result in efficient mixing. In the current version of Dimple, the following built-in factors automatically implement block-Gibbs updates:

- Multinomial
- MultinomialUnnormalizedParameters
- MultinomialEnergyParameters

Most Dimple methods work more-or-less as normal when using the Gibbs solver, but in some cases the interpretation is slightly different than for other solvers. For example, the .Belief method for a discrete variable returns an estimate of the belief based on averaging over the sample values.

NOTE: The setNumIterations() method is not supported by the Gibbs solver as the term “iteration” is ambiguous in this case. Instead, the method setNumSamples() should be used to set the length of the run. The Solver.iterate() method performs a single-variable update in the case of the Gibbs solver, rather than an entire scan of all variables.

The following sections list the solver-specific aspects of the API for the Gibbs solver.

5.6.9.1 Gibbs Options

The following options affect the behavior of various aspects of the Gibbs solver:

5.6.9.1.1 GibbsOptions.numSamples

Type	integer
Default	1
Affects	graph
Description	Specifies the number of samples to be generated when solving the graph post burn-in. (This value times the value of GibbsOptions.numRandomRestarts plus one determines the total number of samples that will be produced.)

5.6.9.1.2 GibbsOptions.scansPerSample

Type	integer
Default	0
Affects	graph
Description	Specifies sampling rate in terms of the number of scans[38] of the graph to perform for each sample.

5.6.9.1.3 GibbsOptions.burnInScans

Type integer
Default 0
Affects graph
Description Specifies the number of scans of the graph to perform during the burn-in phase before generating samples.

5.6.9.1.4 GibbsOptions.numRandomRestarts

Type integer
Default 0
Affects graph
Description Specifies the number of random restarts (zero by default, which means run once and don't restart). For a value greater than zero, the after running the specified number of samples, the solver is restarted with the variable values randomized, and re-run (including burn-in). The sample values (the best sample value, or all samples, if requested) are extracted across all runs.

5.6.9.1.5 GibbsOptions.saveAllSamples

Type boolean
Default false
Affects graph
Description Specifies whether to save all sample values for variables when running Gibbs. Note that this is practical only if the number of variables in the graph times the number of samples per variable is reasonably sized.

5.6.9.1.6 GibbsOptions.saveAllScores

Type boolean
Default false
Affects graph
Description Specifies whether to save scores for all generated samples in Gibbs. If true, then for each sample the total energy/log-likelihood a.k.a. *score* of the graph will be saved. The saved scores can later be retrieved by the getAllScores() method described below.

5.6.9.1.7 GibbsOptions.discreteSampler

Type string
Default CDFSampler
Affects variables
Description Specifies the default sampler to use for discrete variables when a conjugate sampler is not suitable. The sampler may be configured by use of additional options defined by each sampler type. See subsubsection 5.6.13 for more details.

5.6.9.1.8 GibbsOptions.realSampler

Type string
Default SliceSampler
Affects variables
Description Specifies the default sampler to use for real-valued variables (including Complex and RealJoint) when a conjugate sampler is not suitable. The sampler may be configured by use of additional options defined by each sampler type. See subsubsection 5.6.13 for more details.

5.6.9.1.9 GibbsOptions.enableAutomaticConjugateSampling

Type boolean
Default true
Affects variables
Description Specifies whether to use conjugate sampling when available for a given variable. Note that if a specific sampler has been specified for a particular variable (by setting the GibbsOptions.realSampler option directly on the model or solver variable object) then a conjugate sampler will not be used regardless.

5.6.9.1.10 GibbsOptions.computeRealJointBeliefMoments

Type boolean
Default false
Affects variables
Description Specifies whether to compute the belief moments (mean vector and covariance matrix) for RealJoint (and Complex) variables while sampling. To minimize computation, these are not computed by default for RealJoint variables (Real variables always compute similar statistics, and do not have a corresponding option to enable them). If true, the belief moments are computed for each sample on-the-fly (without saving all samples). The computed moments can later be retrieved by the getSampleMean and getSampleCovariance solver-specific methods for the variable (see section 5.6.9.6)

5.6.9.1.11 GibbsOptions.enableAnnealing

Type	boolean
Default	false
Affects	graph
Description	Specifies whether to use a tempering and annealing process when running Gibbs.

5.6.9.1.12 GibbsOptions.annealingHalfLife

Type	double
Default	1.0
Affects	graph
Description	Specifies the rate at which the temperature will be lowered during the tempering and annealing process. This rate is specified in terms of the number of samples required for the temperature to be lowered by half. This value is only used if annealing has been enabled as specified by the enableAnnealing option.

5.6.9.1.13 GibbsOptions.initialTemperature

Type	double
Default	1.0
Affects	graph
Description	Specifies the initial temperature to use when annealing is enabled (as specified by the enableAnnealing option).

5.6.9.2 Graph Methods

The following methods are available on a graph set to use the Gibbs solver:

Variable initialization (both on the first run and subsequent restarts) is randomized whenever possible. For a discrete variable, the value is sampled from the Input (uniform if an input is not specified). For a real variable, if an Input is specified and the Input supports one of the conjugate samplers listed above, that sampler is used to initialize the variable. If bounds are also specified for the variable domain, the values is truncated to fall within the bounds. If only bounds are specified (which are finite above and below), then the value is uniformly sampled from within the bounds. If no finite bounds are specified and there is no input, the variable is initialized to zero (or the value specified by setInitialSampleValue) on the initial run, and left at the final value of the previous run on restart.

Enable or disable the use of tempering, or determine if tempering is in use.

```
graph.Solver.setTemperature(T);
graph.Solver.getTemperature();
```

Set/get the current temperature. Setting the current temperature overrides the current annealing temperature.

```
graph.Solver.getAllScores();
```

Returns an array including the score value for each sample. This method only returns a non-empty value if the GibbsOptions.saveAllScores option was set to true on the graph before generating samples.

```
graph.Solver.getTotalPotential();
```

After running the solver, returns the total potential (score) over all factors of the graph (including input priors on variables) given the most recent sample values.

```
graph.Solver.sample(numSamples)
```

This method runs a specified number of samples without re-initializing, burn-in, or random-restarts (this is distinct from iterate(), which runs a specified number of single-variable updates). Before running this method for the first time, the graph must be initialized using the initialize() method.

```
graph.Solver.burnIn()
```

Run the burn-in samples independently of using solve (which automatically runs the burn-in samples). This may be run before using sample() or iterate().

```
graph.Solver.getRejectionRate()
```

Get the overall rejection rate over the entire graph. The rejection rate is the ratio of the number of MCMC proposals that were rejected to the total number of sampler updates. Rejections only occur in MCMC samplers, such as the MHSampler. In other samplers, such as conjugate samplers or the CDFSampler, rejection doesn't occur and the rate for variables that use these samplers is zero (in these cases, sampling the same value twice in a row is not considered rejection). When getting the rejection rate for the entire graph, both the number of rejections and number of updates is counted for all variables, as well as all blocks of variables over which a block sampler is used. These counts are accumulated from the time the graph is initialized, which automatically occurs when running solve(). This includes both burn-in and subsequent sampling. These counts can also be reset explicitly using the resetRejectionRateStats method (see below), which allows these values to be determined, for example, during a specific set of samples.

```
graph.Solver.resetRejectionRateStats()
```

Explicitly reset the rejection-rate statistics. These are automatically reset when the graph is initialized, which automatically occurs when running solve(), but may be reset manually at other times using this method.

5.6.9.3 Variable Methods

```
variable.Solver.getCurrentSample();
```

Returns the current sample value for a variable.

```
variable.Solver.getAllSamples();
```

Returns an array including all sample values seen so far for a variable. Over multiple variables, samples with the same index correspond to the same joint sample value. This method only returns a non-empty value if the GibbsOptions.saveAllSamples options was enabled for the variable when sampling was performed.

```
variable.Solver.getBestSample();
```

Returns the value of the best sample value seen so far, where best is defined as the sample with the minimum total potential over the graph (sum of -log of the factor values and input priors). When getting the best sample from multiple variables, they all correspond to the same sample in time, thus should be a valid sample from the joint distribution.

```
variable.Solver.setInitialSampleValue(initialSampleValue)
variable.Solver.getInitialSampleValue()
```

Set/get the initial sample value to be used as the starting value for this variable. This value is used only on the first run (not subsequent restarts). Setting this value overrides any randomization of the starting value on the first run.

```
variable.Solver.setSampler(samplerName);
variable.Solver.getSampler();
variable.Solver.getSamplerName();
```

Set/get the sampler to be used for this variable. Setting the sampler for a given variable overrides the default sampler for the given variable type, and also overrides any conjugate sampler that might otherwise be used. Using this method the sampler may be set only to one of the generic samplers appropriate for the given variable type.

The getSampler method returns the sampler object, while the getSamplerName method returns a string indicating the name of the sampler being used for this variable. Automatic assignment of a conjugate sampler is done at graph initialization time, so in order to determine what sampler will actually be used, these methods must be called either after a call to the graph initialize method, or after running the solver.

```
variable.Solver.getRejectionRate()
```

Get the rejection rate for the sampler used for a specific variable. The rejection rate is the ratio of the number of MCMC proposals that were rejected to the total number of sampler updates. Rejections only occur in MCMC samplers, such as the MHSampler. In other samplers, such as conjugate samplers or the CDFSampler, rejection doesn't occur and the rate for variables that use these samplers is zero (in these cases, sampling the same value twice in a row is not considered rejection). These counts are accumulated from the time the graph is initialized, which automatically occurs when running solve(). This includes both burn-in and subsequent sampling. These counts can also be reset explicitly using the resetRejectionRateStats method (see below), which allows these values to be determined, for example, during a specific set of samples.

```
variable.Solver.getNumScoresPerUpdate()
```

Returns the average number of score computations performed per update when sampling from this variable. Use of an MCMC sampler requires computation of the score (energy) for specific settings of the variables. For some samplers, such as the slice sampler, the number of times the score is computed varies, and depends on the particular values and the form of the distribution. The returned value indicates the average number of times the score has been computed. If a non-MCMC-based sampler is used, the returned value will be zero. The count is accumulated from the time the graph is initialized, which automatically occurs when running solve(). This includes both burn-in and subsequent sampling. The count can also be reset explicitly using the resetRejectionRateStats method (see below), which allows this value to be determined, for example, during a specific set of samples.

```
variable.Solver.resetRejectionRateStats()
```

Explicitly reset the rejection-rate statistics for a specific variable (the statistics for computing the rejection rate as well as the number of scores per update). These are automatically reset when the graph is initialized, which automatically occurs when running solve(), but may be reset manually at other times using this method.

5.6.9.4 Discrete-Variable-Specific Methods

The following methods apply only to Discrete, Bit, and FiniteField variables when using the Gibbs solver.

```
variable.Solver.getSampleIndex;
```

Returns the index of the current sample for a variable, where the index refers to the index into the domain of the variable.

```
variable.Solver.getAllSampleIndices;
```

Returns an array including the indices of all samples seen so far for a variable.

```
variable.Solver.getBestSampleIndex;
```

Returns the index of the best sample seen so far.

```
variable.Solver.setInitialSampleIndex(initialSampleIndex)
variable.Solver.getInitialSampleIndex()
```

Set/get the initial sample index associated with the starting value for this variable. The value associated with this index is used only on the first run (not subsequent restarts). Setting this index overrides any randomization of the starting value on the first run.

5.6.9.5 Real-Variable-Specific Methods

The following methods apply only to Real variables when using the Gibbs solver.

```
variable.Solver.getSampleMean;
```

Returns the mean value of all samples that have been collected. This is and estimate of the mean of the belief for the corresponding variable.

```
variable.Solver.getSampleVariance;
```

Returns the variance of all samples that have been collected. This is and estimate of the variance of the belief for the corresponding variable.

5.6.9.6 RealJoint-Variable-Specific Methods

The following methods apply only to RealJoint and Complex variables when using the Gibbs solver.

```
variable.Solver.getSampleMean;
```

Returns the mean vector of all samples that have been collected. This is and estimate of the mean of the belief for the corresponding variable. This method is only available if, prior to performing inference, the option GibbsOptions.computeRealJointBeliefMoments is set to true.

```
variable.Solver.getSampleCovariance;
```

Returns the covariance matrix computed over all samples that have been collected. This is and estimate of the covariance of the belief for the corresponding variable. This method is only available if, prior to performing inference, the option GibbsOptions.computeRealJointBeliefMoments is set to true.

5.6.9.7 Factor Methods

```
factor.Solver.getPotential();
```

Returns the potential value of a factor given the current values of its connected variables.

```
factor.Solver.getPotential(values);
```

Get the potential value of a factor given the variable values specified by the argument vector. The argument must be a vector with length equal to the number of connected variables. For a table-factor (connected exclusively to discrete variables), each value corresponds the index into the domain list for that variable (not the value of the variable itself). For a real-factor (connected to one or more real variables), each value corresponds to the value of the variable.

5.6.9.8 Schedulers and Schedules

The built-in schedulers designed for belief propagation are not appropriate for the Gibbs solver. Instead, there are two built-in schedulers specifically for the Gibbs solver:

- GibbsSequentialScanScheduler
- GibbsRandomScanScheduler

The GibbsSequentialScanScheduler chooses the next variable for updating in a fixed order. It updates all variables in the graph, completing an entire scan, before repeating the same fixed order. (In Gibbs literature this seems to be known as a sequential-scan, systematic-scan, or fixed-scan schedule.)

The GibbsRandomScanScheduler randomly selects a variable for each update (with replacement).

The default scheduler when using the Gibbs solver is the GibbsSequentialScanScheduler, which is used if no scheduler is explicitly specified.

The user may specify a custom schedule when using the Gibbs solver. In this case, the schedule should include only Variable node updates (not specific edges), and no Factor updates (any Factor updates specified will be ignored).

To explicitly specify a scheduler, use the Scheduler or Schedule property of the FactorGraph (see sections 5.1.2.2 and 5.1.2.3).

5.6.9.8.1 Block Schedule Entries

The Gibbs solver allows a schedule to optionally include block entries that allow a group of variables to be updated at once. A block schedule entry can either be included in a custom schedule or added to the schedule produced by one of the Gibbs-specific built-in schedulers. A block schedule entry includes two pieces of information:

- A reference to a block sampler, which is used to perform the update
- A list of variables to be included in this block

In the current version of Dimple, the only built-in block sampler is the BlockMHSampler, which implements block Metropolis-Hastings sampling for the variables included in the block. The BlockMHSampler requires a block proposal kernel to be specified. In the current version of Dimple, there are no built-in general purpose block proposal kernels. To use this sampler, a custom block proposal kernel must be written in Java (see section B.2).

To specify a block schedule entry in a custom schedule, the schedule entry consists of a cell-array in which the first element of the cell is a reference to a Java sampler object, and the subsequent entries are the variables to be included in the block. For example:

```
import com.analog.lyric.dimple.solvers.gibbs.samplers.block.BlockMHSampler;
...
fg.Schedule = {{BlockMHSampler(MyProposalKernel), a, b, c}, d, e};
```

In the above example, we create a block schedule entry that updates variables a, b, and c together, with separate schedule entries for variables d and e. The constructor for the BlockMHSampler requires a proposal kernel. In the above example, “MyProposalKernel” is a user-provided custom proposal class written in Java. (Note that the “import” line in the above example is simply to avoid having to write the fully qualified name each time the BlockMHSampler is used.)

Block schedule entries can also be used with either of the Gibbs-specific built-in schedulers described above. When a block entry is added in this way, for each of the variables included in a block entry, the individual variable entries that would have been present in the schedule are removed. That is, those variables are only included in the corresponding block entry (or entries) and are not also updated independently. In case of the GibbsRandomScanScheduler, each update selects an entry randomly from among all blocks plus all variables that are not in a block.

A block schedule entry can be added when using a built-in Gibbs-specific scheduler using:

```
fg.Scheduler.addBlockScheduleEntry(blockSampler, listOfVariables);
```

The following example shows adding a block schedule entry that includes two elements of the variable x, and variable y.

```
fg.Scheduler.addBlockScheduleEntry(BlockMHSampler(MyKernel), x(2:3), y);
```

Multiple block schedule entries can be added at once using:

```
fg.Scheduler.addBlockScheduleEntries({blockSampler1, listOfVariables1}, {
    blockSampler1, listOfVariables1}, ...);
```

The following example shows adding a block schedule entry that includes two elements of the variable x, and variable y, and a second block that includes variables a and b.

```
fg.Scheduler.addBlockScheduleEntries({BlockMHSampler(MyKernel1), x(2:3), y},
     {BlockMHSampler(MyKernel2), a, b});
```

To implement a custom block proposal kernel, a new Java class must be created that implements the IBlockProposalKernel interface. See section B.2 for more detail.

5.6.10 Particle BP Solver

Use of the particle BP solver is specified by calling:

```
fg.Solver = 'ParticleBP';
```

The following lists the solver-specific options for the ParticleBP solver.

5.6.10.1 Particle BP Options

The following options affect the behavior of various aspects of the ParticleBP solver:

5.6.10.1.1 ParticleBPOptions.numParticles

Type integer

Default 1

Affects real variables

Description Specifies the number of particles used to represent the variable. This option takes affect when the solver variable is constructed and when it is initialized.

5.6.10.1.2 ParticleBPOptions.resamplingUpdatesPerParticle

Type integer

Default 1

Affects real variables

Description For variables on which it is set, specifies the number of updates per particle to perform each time the particle is resampled.

5.6.10.1.3 ParticleBPOptions.iterationsBetweenResampling

Type integer

Default 1

Affects graph

Description Specifies the number of iterations between re-sampling all of the variables in the graph. Default is 1, meaning resample between every iteration.

5.6.10.1.4 ParticleBPOptions.initialParticleRange

Type 1x2 vector

Default [-infinity infinity]

Affects variables

Description Set the range over which the initial particle values will be defined. The initial particle values are uniformly spaced between the min and max values specified. If the range is specified using this method, it overrides any other initial value. Otherwise, if a finite domain has been specified, the initial particle values are uniformly spaced between the lower and upper bound of the domain. Otherwise, all particles are initially set to zero.

```
% Set particle range to the unit interval for all variables in the graph:
graph.setOption('ParticleBPOptions.initialParticleRange', [0.0, 1.0]);
```

5.6.10.1.5 ParticleBPOptions.proposalKernel

Type string

Default NormalProposalKernel

Affects real variables

Description Specifies the type of proposal kernel to use for the specified variables. The selected proposal kernel may have additional options that can be used to configure its behavior. These can also be set on the variables.

5.6.10.1.6 ParticleBPOptions.enableAnnealing

Type boolean

Default false

Affects graph

Description Determines whether to use a tempering and annealing process during inference.

5.6.10.1.7 ParticleBPOptions.annealingHalfLife

Type double

Default 1.0

Affects graph

Description Specifies the rate at which the temperature will be lowered during the tempering and annealing process. This rate is specified in terms of the number of iterations required for the temperature to be lowered by half. This value is only used if annealing has been enabled as specified by the enableAnnealing option.

5.6.10.1.8 ParticleBPOptions.initialTemperature

Type	double
Default	1.0
Affects	graph
Description	Specifies the initial temperature to use when annealing is enabled (as specified by the enableAnnealing option).

5.6.10.2 Graph Methods

The following solver-specific methods are available on the solver graph.

```
graph.Solver.setTemperature(newTemperature);
temp = graph.Solver.getTemperature();
```

Set/get the current temperature. Setting the current temperature overrides the current annealing temperature. This should rarely be necessary.

5.6.10.3 Variable Methods

The Particle BP solver supports both discrete and real variables. For discrete variables, the solver uses sum-product BP as normal, and all of the corresponding methods for the sum-product solver may be used for discrete variables. For real variables, several solver-specific methods are defined, as follows.

5.6.10.4 Real-Variable-Specific Methods

```
variable.Solver.setProposalStandardDeviation(stdDev);
variable.Solver.getProposalStandardDeviation();
```

Set/get the standard deviation for a Gaussian proposal distribution (the default is 1).

```
variable.Solver.getParticleValues();
```

Returns the current set of particle values associated with the variable.

```
variable.Solver.getBelief(valueSet);
```

Given a set of values in the domain of the variable, returns the belief evaluated at these points. The result is normalized relative to the set of points requested so that the sum over the set of returned beliefs is 1.

NOTE: the generic variable method Belief (or getBeliefs() with no arguments) operates similarly to the discrete-variable case, but the belief values returned are those at the current set of particle values. Note that this representation does not represent a set of weighted particles. That is, the particle positions are distributed approximately by the belief and the belief values represent the belief. It remains to be see if this should be the representation of belief that is used, or if an alternative representation would be better. The alternative solver-specific getBelief(valueSet) method allows getting the beliefs on a user-specified set of values, which may be uniform, and would not have this unusual interpretation.

5.6.11 LP Solver

Use of the linear programming (LP) solver is specified by calling:

```
fg.Solver = ’LP’;
```

The LP solver transforms a factor graph MAP estimation problem into an equivalent linear program, which is solved using a linear programming software package. The solver can either be a linear programming solver (in which case the MAP is estimated using an LP relaxation, with no guarantees of correctness), or by an integer linear programming (ILP) solver, in which case the solution is guaranteed to be the MAP. Because this solver release on an external package, you will need to install and configure the specified package before using this solver.

The LP solver supports only discrete variables.

In the current version of Dimple (version 0.7), there is no support for rolled-up graphs when using the LP solver.

The following options are applicable to the LP solver:

5.6.11.1 LP Options

The following options affect the behavior of various aspects of the LP solver:

5.6.11.1.1 LPOptions.LPSolver

Type string
Default ‘’
Affects graph
Description Selects which external LP solver will be used to solve the linear program. Valid values include ‘matlab’, ‘CPLEX’, ‘GLPK’, ‘Gurobi’, ‘LpSolve’, ‘MinSate’, ‘Mosek’, and ‘SAT4J’. The default value is synonomous with specifying ‘matlab’ and will delegate the solver specified by the MatlabLPOption that will be run from the MATLAB frontend. This will obviously only work when running Dimple from MATLAB. None of these solvers are included with Dimple and must be installed and configured separately. The interface for the non-MATLAB based solvers is provided by the third-party Java ILP package. See javailp.sourceforge.net for more information about configuring various solvers.

5.6.11.1.2 LPSolver.MatlabLPSolver

Type string

Default ‘’

Affects graph

Description Selects which LP solver will be run from MATLAB to solve the linear program. This option is only relevant if the LPSolver option has been set to ‘matlab’ and Dimple is being run from MATLAB. The choices for the string solvername are ‘matlab’, ‘glpk’, ‘glpkIP’, ‘gurobi’, and ‘gurobiIP’. The ‘matlab’, ‘glpk’, and ‘gurobi’ solvers are linear programming solvers, while ‘glpkIP’ and ‘gurobiIP’ are ILP solvers. The default value is synonomous with ‘matlab’.

Using the matlab LP solver requires the the MATLAB Optimization Toolbox. Using ‘glpk’ or ‘glpkIP’ requires glpkmex to be in the matlab path, and ‘gurobi’ and ‘gurobiIP’ require the gurobi matlab interface to be in the matlab path; in either case the appropriate packages will need to be obtained and installed.

5.6.12 Proposal Kernels

The following proposal kernels are provided by Dimple for use in samplers. Additional kernels may be added by creating new proposal kernel Java classes that implement the appropriate interfaces.

5.6.12.1 NormalProposalKernel

The following options may be used to configure this kernel:

5.6.12.1.1 NormalProposalKernel.standardDeviation

Type	double
Default	1.0
Affects	variables
Description	Specifies the standard deviation to use on NormalProposalKernel instances attached to the variables that are affected by the option setting. The value must be non-negative.

5.6.12.2 CircularNormalProposalKernel

The CircularNormalProposalKernel makes proposals from a Normal distribution on a circularly wrapping range of the real line. For example, setting the bounds of the range to $-\pi$ and π would create proposals representing angles on a circle.

Since this is a subclass of NormalProposalKernel, the standardDeviation option defined for that class will also affect this one. The following additional options may also be used:

5.6.12.2.1 CircularNormalProposalKernel.lowerBound

Type	double
Default	$-\pi$
Affects	variables
Description	Specifies lower bound to use on CircularNormalProposalKernel instances attached to the variables that are affected by the option setting.

5.6.12.2.2 CircularNormalProposalKernel.upperBound

Type	double
Default	$+\pi$
Affects	variables
Description	Specifies upper bound to use on CircularNormalProposalKernel instances attached to the variables that are affected by the option setting.

5.6.12.3 UniformDiscreteProposalKernel

This kernel does not have any configurable options.

5.6.13 Samplers

Samplers are used by sampling solvers (e.g. Gibbs) to generate samples for variables either singly or in blocks.

The following non-conjugate single variable samplers are currently available:

5.6.13.1 CDFSampler

The CDFSampler can be used with discrete variable types. It samples from the full conditional distribution of the variable. It is the default sampler used for discrete variables in the Gibbs solver.

It does not support any configuration options.

5.6.13.2 MHSampler

The MHSampler may be used for discrete or real variables. It implements the Metropolis-Hastings sampling algorithm. The sampler may be configured to use different proposal kernels for discrete and real variables as described below.

The following options may be used to configure the MHSampler for a given variable:

5.6.13.2.1 MHSampler.discreteProposalKernel

Type	string
Default	UniformDiscreteProposalKernel
Affects	variables
Description	Specifies the proposal kernel to use when using the MHSampler on discrete variables for which this option setting is visible. Depending on the kernel selected, it may also be configured by additional option settings. See subsubsection 5.6.12 for more details.

5.6.13.2.2 MHSampler.realProposalKernel

Type string

Default NormalProposalKernel

Affects variables

Description Specifies the proposal kernel to use when using the MHSampler on real variables for which this option setting is visible. Depending on the kernel selected, it may also be configured by additional option settings. See subsubsection 5.6.12 for more details.

5.6.13.3 SliceSampler

The slice sampler may be used with real variables. It implements the algorithm described in Neal's paper "Slice Sampling" 2010. It is the default sampler used for real variables in the Gibbs solver.

The following options may be used to configure the SliceSampler for a given variable:

5.6.13.3.1 SliceSampler.initialSliceWidth

Type double

Default 1.0

Affects variables

Description The size of the initial slice to use when sampling for SliceSampler instances used by variables that are affected by the option setting. For variables with a natural range that is much smaller or much larger than the default value of one, it may be beneficial to modify this value.

5.6.13.3.2 SliceSampler.maximumDoublings

Type integer

Default 10

Affects variables

Description The maximum number of doublings used during the doubling phase of the slice sampler. The maximum slice interval is on the order of $initialSliceWidth \cdot 2^{maximumDoublings}$

5.6.13.4 SuwaTodoSampler

The Suwo-Todo sampler can be used for discrete variables. It implements the algorithm described in Suwa and Todo's paper "Markov Chain Monte Carlo Method without Detailed Balance" (2010).

This sampler does not support any configuration options.

5.7 Streaming Data

5.7.1 Variable Stream Common Properties and Methods

Dimple supports several types of variable streams:

- DiscreteStream
- BitStream
- RealStream
- RealJointStream
- ComplexStream

Each of these share common properties and method listed in the following sections.

5.7.1.1 Properties

5.7.1.1.1 DataSource

Read-write. When written, connects the variable stream to a data source (see section 5.7.8). The data sink must be of a type appropriate for the particular variable stream type.

5.7.1.1.2 DataSink

Read-write. When written, connects the variable stream to a data sink (see section 5.7.12). The data sink must be of a type appropriate for the particular variable stream type.

5.7.1.1.3 Dimensions

Read-only. Indicates the dimensions of the variable stream. The dimensions correspond to the size of the variable array at each position in the stream.

5.7.1.1.4 Size

Read-only. Indicates the number of elements in the variable stream that are actually instantiated. Each element corresponds to one copy of the variable or variable array at a specific point in the stream. Dimple instantiates the minimum number of contiguous elements to cover the slices of the stream that are actually used in factors (see section 5.7.1.2.1), plus the number of additional elements to cover the indicated BufferSize (see section 5.7.7.1.1).

5.7.1.1.5 Variables

Read-only. Returns a variable array containing all of the currently instantiated variables in the stream.

5.7.1.1.6 Domain

Read-only. Returns the domain in a form that depends on the variable type, as summarized in the following table:

Stream Type	Domain Data Type
DiscreteStream	DiscreteDomain (see section 5.2.9)
BitStream	DiscreteDomain (see section 5.2.9)
RealStream	RealDomain (see section 5.2.10)
RealJointStream	RealJointDomain (see section 5.2.11)
ComplexStream	ComplexDomain (see section 5.2.12)

5.7.1.2 Methods

5.7.1.2.1 getSlice

```
varStream.getSlice(startIndex);
```

The getSlice method is used to extract a *slice* of the stream, which means a version of the stream that may be offset from the original stream itself. This is generally used for specifying streams to connect to a factor when calling addFactor.

Takes a single numeric argument, startIndex, which indicates the starting position in the stream. The resulting stream slice is essentially a reference to the stream offset by startIndex-1. For example, a startIndex of 2 returns a slice offset by 1, such that the first location in the slice corresponds to the second location in the original stream. A startIndex of 1 returns a slice identical to the original stream.

5.7.2 DiscreteStream

5.7.2.1 Constructor

The DiscreteStream constructor is used to create a stream of Discrete variables or arrays of Discrete variables.

```
DiscreteStream(domain, [dimensions]);
```

- domain is a required argument indicating the domain of the variable. The domain may either be a numeric array of domain elements, a cell array of domain elements, or a DiscreteDomain object (see section 5.2.3.1.1).
- dimensions is an optional variable-length comma-separated list of matrix dimensions (an empty list indicates a single Discrete variable).

5.7.3 BitStream

5.7.3.1 Constructor

The BitStream constructor is used to create a stream of Bit variables or arrays of Bit variables.

```
BitStream([dimensions]);
```

- dimensions is an optional variable-length comma-separated list of matrix dimensions (an empty list indicates a single Bit variable stream).

5.7.4 RealStream

5.7.4.1 Constructor

The RealStream constructor is used to create a stream of Real variables or arrays of Real variables.

```
RealStream([domain], [dimensions]);
```

All arguments are optional and can be used in any combination.

- domain specifies a bound on the domain of the stream. It can either be specified as a two-element array or a RealDomain object (see section 5.2.10). If specified as an array, the first element is the lower bound and the second element is the upper bound. -Inf and Inf are allowed values for the lower or upper bound, respectively. If no domain is specified, then a domain from $-\infty$ to ∞ is assumed.
- dimensions specify the array dimensions. The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

5.7.5 RealJointStream

5.7.5.1 Constructor

The RealJointStream constructor is used to create a stream of RealJoint variablesor arrays of RealJoint variables.

```
RealJointStream(domain, [dimensions])
RealJointStream(numJointVariables, [dimensions]);
```

The arguments are defined as follows:

- domain specifies a bound on the domain of the variable. It is specified by a RealJointDomain object (see section 5.2.11). If no domain is specified, then an unbounded domain is assumed and numJointVariables must be specified instead.
- numJointVariables specifies the number of joint real-valued elements. This argument is present only if the domain argument is not specified.
- dimensions specify the array dimensions (the array of individual RealJointStream variables). The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

5.7.6 ComplexStream

5.7.6.1 Constructor

The ComplexStream constructor is used to create a stream of Complex variablesor arrays of Complex variables .

```
ComplexStream([domain], [dimensions]);
```

The arguments are defined as follows:

- domain specifies the domain of the ComplexStream using a ComplexDomain object (see 5.2.12). If no domain is specified, then an unbounded domain is assumed.
- dimensions specify the array dimensions (the array of individual Complex variables). The behavior of the list of dimensions is identical to that for Discrete variables as described in section 5.2.3.1.2.

5.7.7 FactorGraphStream

A FactorGraphStream is constructed automatically and returned as the result of adding a factor to a graph using the addFactor method where one or more of the arguments are variable streams.

5.7.7.1 Properties

5.7.7.1.1 BufferSize

Read-write. When written, modifies the number of instantiated elements in the FactorGraphStream to include the specified number of copies of the corresponding factor and connected variables. By default, the BufferSize is 1. When running the solver on one step of the overall factor graph, the solver uses the entire buffer. Making the buffer size larger means using more of the history in performing inference for each step. The results of inference run on previous steps that is beyond the size of the buffer is essentially frozen, and is no longer updated on subsequent steps of the solver.

5.7.8 Data Source Common Properties and Methods

Dimple supports several types of streaming data sources. A data source is a source of Input values to the variables within a variable stream[39]. When performing inference, as each step that the graph is advanced, the next source value is read from the data source, and the earlier values are shifted back to earlier time-steps in the graph.

Each source type corresponds to a particular format of input data. Each type is appropriate only to a specific type of variable stream and solver. The following table summarizes these requirements.

Data Source	Variable Stream Type	Supported Solvers
DoubleArrayDataSource	DiscreteStream, BitStream	all
	RealStream	SumProduct
MultivariateDataSource	RealJointStream	SumProduct
FactorFunctionDataSource	RealStream	SumProduct, Gibbs, ParticleBP

[39] In the current version of Dimple, data sources are limited to providing Inputs to variables. A future version of Dimple may expand this capability to allow sourcing FixedValues or other types of input data.

Each of these share common properties and method listed in the following sections.

5.7.8.1 Properties

5.7.8.1.1 Dimensions

Read-only. Indicates the dimensions of the data source. The dimensions correspond to the size of the variable array at each position in the stream that the data source will feed.

5.7.9 DoubleArrayDataSource

5.7.9.1 Constructor

```
DoubleArrayDataSource([dimensions], [initialData]);
```

- dimensions is an optional row vector indicating the matrix dimensions of the data source (an empty argument indicates a single data source to connect to a single variable stream).
- initialData is an optional array of data the stream will contain (more data can be added later). This is a multidimensional array where the first dimensions correspond to the dimension of the variable stream this will feed, the next dimensions corresponds to the length of the Input vector for each variable (the domain size for discrete variable streams, and 2 for real variable streams used with the SumProduct solver), and the final dimension is the number of time-steps of data to provide. For single variable streams, the first dimensions are omitted.

5.7.9.2 Methods

5.7.9.2.1 add

```
dataSource.add(data);
```

This method appends the data source with the specified data. The data argument is a multidimensional array, where the first dimensions correspond to the dimension of the variable stream this will feed, the next dimensions corresponds to the length of the Input vector for each variable (the domain size for discrete variable streams, and 2 for real variable streams used with the SumProduct solver), and the final dimension is the number of time-steps of data to provide. For single variable streams, the first dimensions are omitted.

5.7.10 MultivariateDataSource

5.7.10.1 Constructor

```
MultivariateDataSource([dimensions]);
```

- dimensions is an optional row vector indicating the matrix dimensions of the data source (an empty argument indicates a single data source to connect to a single variable stream).

5.7.10.2 Methods

5.7.10.2.1 add

```
dataSource.add(means, covariances);
```

This method appends the data source with the a single time-step of data (multiple time-steps must be added using successive calls to this method). Means should have dimensions [VariableDimensions NumberJointVariables] and Covariances should be of the form [VariableDimensions NumberJointVariables NumberJointVariables].

5.7.11 FactorFunctionDataSource

5.7.11.1 Constructor

```
FactorFunctionDataSource([dimensions]);
```

- dimensions is an optional row vector indicating the matrix dimensions of the data source (an empty argument indicates a single data source to connect to a single variable stream).

5.7.11.2 Methods

5.7.11.2.1 add

```
dataSource.add(data);
```

This method appends the data source with the a single time-step of data (multiple time-steps must be added using successive calls to this method).The data argument is a multidimensional cell array, with dimension equal to the corresponding dimensions of the variable stream this will feed. Each element is a FactorFunctions (see section 5.3.3), which is to represent the Input of the corresponding variable.

5.7.12 Data Sink Common Properties and Methods

Dimple supports several types of streaming data sinks:

Dimple supports several types of streaming data sinks. A data sink is a data structure used to store successive results of inference from the variables with a variable stream. Specifically, it stores the Belief values of these variables[40]. When performing inference, as each step that the graph is advanced, the Belief value for the earliest element of the variable stream is stored in the data sink.

Each sink type corresponds to a particular format of output data. Each type is appropriate only to a specific type of variable stream and solver. The following table summarizes these requirements.

Data Sink	Variable Stream Type	Supported Solvers
DoubleArrayDataSink	DiscreteStream, BitStream	all
	RealStream	SumProduct
MultivariateDataSink	RealJointStream	SumProduct

Each of these share common properties and method listed in the following sections.

[40] In the current version of Dimple, data sinks are limited to the Beliefs to variables. A future version of Dimple may expand this capability to allow sinking other types of result data.

5.7.12.1 Properties

5.7.12.1.1 Dimensions

Read-only. Indicates the dimensions of the data sink. The dimensions correspond to the size of the variable array at each position in the stream that the data sink will be fed from.

5.7.12.2 Methods

5.7.12.2.1 hasNext

```
hasNext = dataSink.hasNext();
```

Used in connection with the getNext method (described in the sections below), this method takes no arguments and returns a boolean indicating whether or not there are any more time steps in the dataSink that have not yet been extracted.

5.7.13 DoubleArrayDataSink

5.7.13.1 Constructor

```
DoubleArrayDataSink([dimensions]);
```

- dimensions is an optional array indicating the matrix dimensions of the data sink (an empty argument indicates a single data sink to connect to a single variable stream).

5.7.13.2 Properties

5.7.13.2.1 Array

Read-only. Extracts the entire contents of the data sink as an array. The first dimensions of the array correspond to the Dimensions of the data sink , the next dimension corresponds to the size of the belief array for each variable,and the final dimension corresponds to the number of time steps that had been gathered. For discrete variables, the dimension of the belief array corresponds to the domain sizes, while for real variables used with the SumProduct solver the dimension is 2, where the elements correspond to the mean and standard deviation, respectively.

5.7.13.3 Methods

5.7.13.3.1 getNext

```
b = dataSink.getNext();
```

This method takes no arguments, and returns the set of belief values from the next time-step. The returned value is a multidimensional array, where the first dimensions correspond to the Dimension of the data sink, and the next dimension corresponds to the size of the belief array for each variable. For discrete variables, the dimension of the belief array corresponds to the domain sizes, while for real variables used with the SumProduct solver the dimension is 2, where the elements correspond to the mean and standard deviation, respectively.

5.7.14 MultivariateDataSink

5.7.14.1 Constructor

```
MultivariateDataSink([dimensions]);
```

- dimensions is an optional array indicating the matrix dimensions of the data sink (an empty argument indicates a single data sink to connect to a single variable stream).

5.7.14.2 Methods

5.7.14.2.1 getNext

```
b = dataSink.getNext();
```

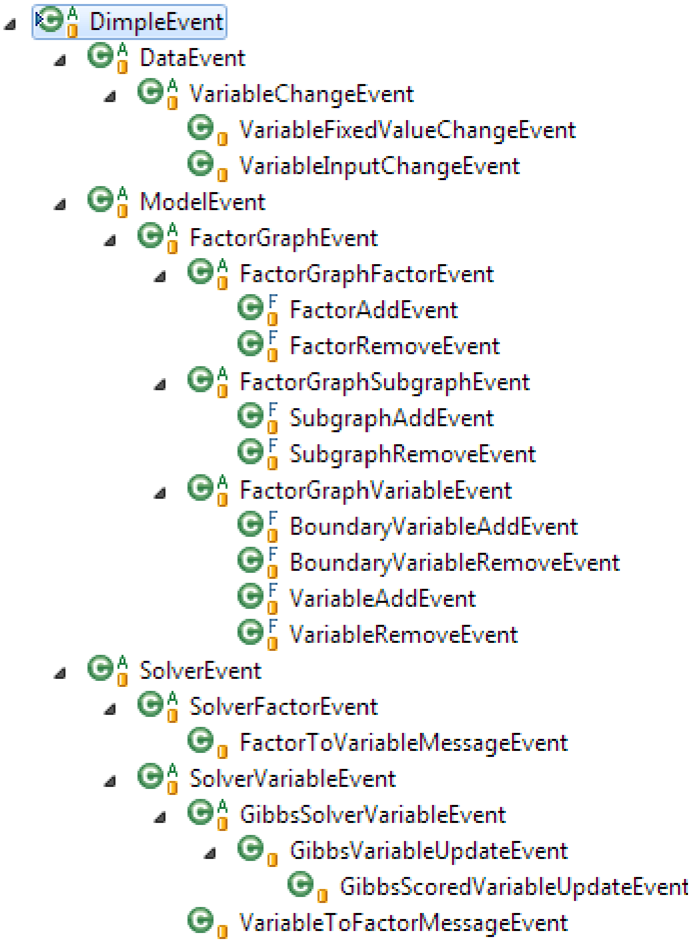
This method takes no arguments, and returns the set of belief values from the next timestep. The returned value is a cell array of MultivariateNormalParameters objects (see section 5.2.15), each of which contains and mean vector and covariance matrix. The dimensions of this cell array correspond to the Dimension of the data sink.

5.8 Event Monitoring

Sometimes it can be useful to monitor the actions Dimple takes as the model or data changes or as inference is performed. Such monitoring can be helpful when debugging your model, when trying to determine whether inference has converged while using belief propogation on a loopy graph, or when attempting to determine whether a graph has been adequately mixed when using the Gibbs solver. To address this need, Dimple provides an event-based system consisting of events that can be triggered when various actions of interest occur and associated event handlers and an event listener that handles dispatching of events. The event system is designed to have no effect on the performance of inference when it has not been enabled, but may have a noticeable effect when it is being used.

The full power of the event system is only available directly in the Java API but the MATLAB API does provide a simple event logging interface that allows events to be logged to the console or an external log file.

5.8.1 Event types



```
DimpleEvent
    DataEvent
        VariableChangeEvent
            VariableFixedValueChangeEvent
            VariableInputChangeEvent
    ModelEvent
        FactorGraphEvent
            FactorGraphFactorEvent
                FactorAddEvent
                FactorRemoveEvent
            FactorGraphSubgraphEvent
                SubgraphAddEvent
                SubgraphRemoveEvent
            FactorGraphVariableEvent
                BoundaryVariableAddEvent
                BoundaryVariableRemoveEvent
                VariableAddEvent
                VariableRemoveEvent
    SolverEvent
        SolverFactorEvent
            FactorToVariableMessageEvent
        SolverVariableEvent
            GibbsSolverVariableEvent
                GibbsVariableUpdateEvent
                    GibbsScoredVariableUpdateEvent
            VariableToFactorMessageEvent
```

Figure 1: Dimple event hierarchy

Dimple Events are organized hierarchically, and the current version of the full hierarchy is shown in Figure 1. Note that any event types marked with an 'A' in the diagram are abstract super types and are used to organize the events by category. Actual event instances will belong to non-abstract types, which are for the most part leaf-nodes in the diagram. Events are subdivided into three categories:

- ModelEvent: includes changes to the structure of the model including adding and removing variables, factors and subgraphs. The concrete model event types are:
    - FactorAddEvent and FactorRemoveEvent: raised when a factor is added or removed from the model.
    - VariableAddEvent and VariableRemoveEvent: raised when a variable is added or removed from the model.
    - SubgraphAddEvent and SubgraphRemoveEven: raised when a subgraph is added or removed from the model.

  - BoundaryVariableAddEvent and BoundaryVariableRemoveEvent: raised when a boundary variable is added or removed.

- DataEvent: includes changes to data including changes to fixed values and inputs. The concrete data event types are:

  - VariableFixedValueChangeEvent: raised when a fixed value is set or unset on a variable.
  - VariableInputChangeEvent: raised when an input distribution is set or unset on a variable.

- SolverEvent: includes solver-specific events of interest that occur while running inference. The concrete event types are:

  - FactorToVariableMessageEvent and VariableToFactorMessageEvent: raised when edge messages are updated in belief propagation solvers. Currently only the sumproduct and minsum solvers generate these messages.
  - GibbsVariableUpdateEvent and GibbsScoredVariableUpdateEvent: raised when sample values are changed by the Gibbs solver. The two event types are the same except that the scored version adds information about the change in score induced by the sample update.

Additional events may be added in future releases. New event types may also be added by developers who have extended Dimple with their own solvers or custom factors in Java.

When specifying event types in the MATLAB API, use the event type name in a string value:

```
logger = getEventLogger();
logger.log('FactorToVariableMessageEvent', fg);
```

Unrecognized event types will result in an exception.

Note that Dimple events are not MATLAB events and you cannot use MATLAB's event API to trigger or listen for Dimple events.

5.8.2 Event logging

The easiest way to monitor events in Dimple is through an event logger. Given an event logger instance, you can configure it to log either to the console or to a text file, you can configure how verbose the output should be, and specify which events for which model objects should be logged.

Event loggers are instances of the EventLogger class. You may create an instance using the constructor:

```
logger = EventLogger();
```

or you may simply get the default global instance using the getEventLogger() function, which will create a new logger the first time it is invoked:

```
logger = getEventLogger();
```

Newly constructed loggers will output to standard error by default and will have a default verbosity of zero, which will produce the most terse output. You may configure the logger to change the verbosity level or to direct output to a different target. For example:

```
% Use more verbose log output.
logger.Verbosity = 2;

% Append output to a file in the working directory.
logger.open('event-logger.txt');

% ... or output to standard output
logger.open('stdout');
```

Usually a single logger will be sufficient, but you can create multiple logger objects that direct output to different targets.

To enable logging for a particular class of events on your model, use the log method with the event type of interest. If the event type is abstract (is annotated with the letter A in Figure 1), then all event subtypes will be logged. If the event type is not abstract, then only that particular event type will be logged. In particular, if you specify SolverEvent on a graph using the Gibbs solver, you will see GibbsScoredVariableUpdateEvent messages but if you specify GibbsVariableUpdateEvent you will get only messages for that specific class and will not get scored messages.

```
% Log all solver events for given model.
logger.log('SolverEvent', model);

% Log unscored Gibbs update messages
logger.log('GibbsVariableUpdateEvent', model);

% Log variable to factor messages for a single variable
logger.log('VariableToFactorMessageEvent', x);
```

You may remove previously created log registration either by using the unlog method to remove individual entries or the clear method to remove all entries. When using unlog, the arguments must match the original arguments passed to log. (Note that setting the verbosity to a negative value or closing the output will also turn off logging it will not prevent Dimple from creating the underlying events, so make sure to use the clear() method when you are done with logging if you do not want to slow down inference.)

```
% Disable a previous log registration
logger.log('SolverEvent', model);

% Disable all logging
logger.clear();
```

When using logging, it is usually very helpful to give the variables, factors and subgraphs unique, easy to read, names. This will make your log output much easier to understand.

5.8.3 Advanced event handling

The MATLAB API does not support handling Dimple events outside of the EventLogger class. More advanced uses cases may be implemented using the Java API. See the Java version of this manual for details.

5.9 List of Built-in Factors

The following table lists the current set of built-in Dimple factors. For each, the name is given, followed by the set of variables that would be connected to the factor, followed by any constructor arguments. Optional variables and constructor arguments are in brackets. And an arbitrary length list or array of variables is followed by ellipses. The allowed set of variable data-types for each variable is given in parentheses (B = Bit, D = Discrete, F = FiniteFieldVariable, R = Real, C = Complex, RJ = RealJoint).

Name	Variables	Constructor	Description

Abs	out(D,R)
	in(D,R)
		[smoothing]	Deterministic absolute value function, where out = abs(in). An optional smoothing value may be specified as a constructor argument[41].

ACos	out(D,R)
	in(D,R)
		[smoothing]	Deterministic arc-cosine function, where out = acos(in). An optional smoothing value may be specified as a constructor argument[41].

AdditiveNoise	out(R)
	in(B,D,R)
		σ	Add Gaussian noise with a known standard deviation, σ, specified in constructor.

And	out(B)
	in...(B)
		-	Deterministic logical AND function, where out = AND(in...).

ASin	out(D,R)
	in(D,R)
		[smoothing]	Deterministic arc-sine function, where out = asin(in). An optional smoothing value may be specified as a constructor argument[41].

ATan	out(D,R)
	in(D,R)
		[smoothing]	Deterministic arc-tangent function, where out = atan(in). An optional smoothing value may be specified as a constructor argument[41].

Bernoulli	[ρ](R) x...(B)	[ρ]	Bernoulli distribution, $p(x\|\rho)$, where ρ is a parameter indicating the probability of one, and x is an array of Bit variables. There can be any number of x variables, all associated with the same parameter value. The conjugate prior for the parameter, ρ, is a Beta distribution[42]. The parameter, ρ, can be a variable or a constant specified in the constructor.

[41] If smoothing is enabled, the factor function becomes $e^{-(\text{out}-F(\text{in}))^2/\text{smoothing}}$ (making it non-deterministic) instead of $\delta(\text{out} - F(\text{in}))$, where F is the deterministic function associated with this factor. This is useful for solvers that do not work well with deterministic real-valued factors, such as particle BP, particularly when annealing is used.

[42] It is not necessary to use the conjugate prior, but in some cases there may be a benefit.

Name	Variables	Constructor	Description
Beta	$[\alpha]$(R) $[\beta]$(R) value...(R)	$[\alpha]$ $[\beta]$	Beta distribution. There can be any number of value variables, all associated with the same parameter values. Parameters α and β can be variables, or if both are constant they can be specified in the constructor.
Binomial	$[N]$(D) ρ(R) x(D)	$[N]$	Binomial distribution, $p(x\|N, \rho)$, where N is the total number of trials, ρ is a parameter indicating the success probability, and x is a count of success outcomes. Parameter N can be a Discrete variable with positive integer values or a constant integer value specified in the constructor. The domain of x must include integers from 0 through N, or if N is a variable, through the maximum value in the domain of N. The conjugate prior for the parameter, ρ, is a Beta distribution[42].
Categorical	$[\alpha]$(RJ) x...(D)	$[\alpha]$	Categorical distribution, $p(x\|\alpha)$, where α is a vector of parameter variables and x is an array of discrete variables. The number of elements in α must equal the domain size of x. There can be any number of x variables, all associated with the same parameter values. The α parameters are represented as a normalized probability vector. The conjugate prior for this representation is a Dirichlet distribution[42]. In the current implementation, the domain of the x variable must be zero-based contiguous integers, $0...N-1$[43].

[43] This limitation may be lifted in a future version.

Name	Variables	Constructor	Description
Categorical EnergyParameters	$[\alpha]$...(R) x...(D)	N, $[\alpha]$	Categorical distribution, $p(x\|\alpha)$, where α is a vector of parameter variables and x is an array of discrete variables. The number of elements in α and the domain size of x must equal the value of the constructor argument, N. There can be any number of x variables, all associated with the same parameter values. In this alternative version of the Categorical distribution, the α parameters are represented as energy values, that is, $\alpha = -\log(\rho)$, where ρ are unnormalized probabilities. The conjugate prior for this representation is such that each entry of α is independently distributed according to a negative exp-Gamma distribution, all with a common β parameter[42]. In the current implementation, the domain of the x variable must be zero-based contiguous integers, $0...N-1$[43].
Categorical Unnormalized Parameters	$[\alpha]$...(R) x...(D)	N, $[\alpha]$	Categorical distribution, $p(x\|\alpha)$, where α is a vector of parameter variables and x is an array of discrete variables. The number of elements in α and the domain size of x must equal the value of the constructor argument, N. There can be any number of x variables, all associated with the same parameter values. In this alternative version of the Categorical distribution, the α parameters are represented as a vector of unnormalized probability values. The conjugate prior for this representation is such that each entry of α is independently distributed according to a Gamma distribution, all with a common β parameter[42]. In the current implementation, the domain of the x variable must be zero-based contiguous integers, $0...N-1$[43].

ComplexAbs	out(R)
	in(C)
		[smoothing]	Deterministic complex absolute value, where out = $\sqrt{Re(\text{in}) + Im(\text{in})}$. An optional smoothing value may be specified as a constructor argument[41].

ComplexConjugate	out(C)
	in(C,R)
		[smoothing]	Deterministic complex conjugate function, where out = in*. An optional smoothing value may be specified as a constructor argument[41].

Name	Variables	Constructor	Description

ComplexDivide	quotient(C)
	dividend(C,R)
	divisor(C,R)
		[smoothing]	Deterministic complex divide function, where quotient = $\frac{\text{dividend}}{\text{divisor}}$. An optional smoothing value may be specified as a constructor argument^{41}.

ComplexExp	out(C)
	in(C,R)
		[smoothing]	Deterministic complex exponentiation function, where out = exp(in). An optional smoothing value may be specified as a constructor argument^{41}.

ComplexNegate	out(C)
	in(C,R)
		[smoothing]	Deterministic complex negation function, where out = -in. An optional smoothing value may be specified as a constructor argument^{41}.

ComplexProduct	out(C)
	in...(C,R)
		[smoothing]	Deterministic complex product function, where out = $\prod$ in. An optional smoothing value may be specified as a constructor argument^{41}.

ComplexSubtract	out(C)
	posIn(C,R)
	negIn...(C,R)
		[smoothing]	Deterministic complex subtraction function, where out = posIn $-\sum$ negIn. An optional smoothing value may be specified as a constructor argument^{41}.

ComplexSum	out(C)
	in...(C,R)
		[smoothing]	Deterministic complex summation function, where out = $\sum$ in. An optional smoothing value may be specified as a constructor argument^{41}.

ComplexTo
RealAndImaginary
	outReal(R)
	outImag(R)
	in(RJ)
		[smoothing]	Deterministic conversion of a Complex variable to two Real variables, with the first representing the real component and the second representing the imaginary component. An optional smoothing value may be specified as a constructor argument^{41}.

Cos	out(D,R)
	in(D,R)
		[smoothing]	Deterministic cosine function, where out = cos(in). An optional smoothing value may be specified as a constructor argument^{41}.

Cosh	out(D,R)
	in(D,R)
		[smoothing]	Deterministic hyperbolic-cosine function, where out = cosh(in). An optional smoothing value may be specified as a constructor argument^{41}.

Dirichlet	[α](RJ) value...(RJ)	[α]	Dirichlet distribution. There can be any number of value variables, all associated with the same parameter values. Parameter vector α can be a RealJoint variable or a constant specified in the constructor. The dimension of α and each of the value variables must be identical.

Name	Variables	Constructor	Description

DiscreteTransition	y(D)
	x(D)
	A...(RJ)
		-	Parameterized discrete transition factor, $p(y\|x, A)$, where x and y are discrete variables, and A is a matrix of transition probabilities. The transition matrix is organized such that columns correspond to the output distribution for each input state. That is, the transition matrix multiplies on the left. Each column of A corresponds to a RealJoint variable. The number of columns in A must equal the domain size of x, and the dimension of each element of A must equal the domain size of y.
			Each element of A corresponds to a normalized probability vector. The conjugate prior for this representation is such that each element of A is distributed according to a Dirichlet distribution[42].
			In the current implementation, the domain of the x variable must be zero-based contiguous integers, $0...N-1$[43].

DiscreteTransition
EnergyParameters
	y(D)
	x(D)
	A...(R)
		$N_y, N_x\|$
		N
			Parameterized discrete transition factor, $p(y\|x, A)$, where x and y are discrete variables, and A is a matrix of transition probabilities. The transition matrix is organized such that columns correspond to the output distribution for each input state. That is, the transition matrix multiplies on the left. The number of columns in A and the domain size of x must equal the value of the constructor argument, N_x and the number of rows in A and the domain size of y must equal the value of the constructor argument N_y. If N_x and N_y are equal, a single constructor argument, N, may be used.
			The elements of the matrix A are represented as energy values, that is, $A_{i,j} = -\log(\rho_{i,j})$, where ρ is an unnormalized transition probability matrix. The conjugate prior for this representation is such that each entry of A is independently distributed according to a negative exp-Gamma distribution, all with a common β parameter[42].
			In the current implementation, the domain of the x variable must be zero-based contiguous integers, $0...N-1$[43].

Name	Variables	Constructor	Description

DiscreteTransition
Unnormalized
Parameters
	y(D)
	x(D)
	A...(R)
		$N_y, N_x\|$
		N
			Parameterized discrete transition factor, $p(y\|x, A)$, where x and y are discrete variables, and A is a matrix of transition probabilities. The transition matrix is organized such that columns correspond to the output distribution for each input state. That is, the transition matrix multiplies on the left. The number of columns in A and the domain size of x must equal the value of the constructor argument, N_x and the number of rows in A and the domain size of y must equal the value of the constructor argument N_y. If N_x and N_y are equal, a single constructor argument, N, may be used.
			The elements of the matrix A are represented as unnormalized probability values. The conjugate prior for this representation is such that each entry of A is independently distributed according to a Gamma distribution, all with a common β parameter[42].
			In the current implementation, the domain of the x variable must be zero-based contiguous integers, $0...N-1$[43].

Divide	quotient(D,R)
	dividend(D,R)
	divisor(D,R)
		[smoothing]	Deterministic divide function, where quotient $= \frac{\text{dividend}}{\text{divisor}}$. An optional smoothing value may be specified as a constructor argument[41].

Equality	value...(B,D,R)	[smoothing]	Deterministic equality constraint. An optional smoothing value may be specified as a constructor argument[41].

Equals	out(B)
	in...(B,D,R)
		-	Deterministic equals function, where out = (in(1) == in(2) == ...).

ExchangeableDirichlet	$[\alpha]$(R) value...(RJ)	N, $[\alpha]$	Exchangeable Dirichlet distribution. This is a variant of the Dirichlet distribution parameterized with a single common parameter for all dimensions. There can be any number of value variables, all associated with the same parameter value. Parameter α can be a Real variable or a constant specified in the constructor. The dimension of each of the value variables must be identical and equal to the value of N, specified in the constructor.

Name	Variables	Constructor	Description

Exp	out(D,R)
	in(D,R)
		[smoothing]	Deterministic exponentiation function, where out = exp(in). An optional smoothing value may be specified as a constructor argument[41].

FiniteFieldAdd	out(F)
	in1(F)
	in2(F)
		-	Deterministic finite field two-input addition. See section 4.4 for a description of how to use finite field variables.

FiniteFieldMult	out(F)
	in1(F)
	in2(F)
		-	Deterministic finite field two-input multiplication. See section 4.4 for a description of how to use finite field variables.

FiniteFieldProjection	fieldVar(F)
	indices(const)
	bits...(B)
		-	Deterministic projection of a finite field variable onto a set of bit variables corresponding to the bits of the field value. The indices argument is a constant array, which must be a permutation of 0 through $N-1$, where N is the number of bits in the finite field value. The indices represent the order of the projection of the bits in the finite field value onto the corresponding Bit variable in the list of bits. See section 4.4 for a description of how to use finite field variables.

Gamma	[α](R) [β](R) value...(R)	[α] [β]	Gamma distribution. There can be any number of value variables, all associated with the same parameter values. Parameters α and β can be variables, or if both are constant they can be specified in the constructor.

GreaterThan	out(B)
	in1(B,D,R)
	in2(B,D,R)
		-	Deterministic greater-than function, where out = in1 $>$ in2.

InverseGamma	[α](R) [β](R) value...(R)	[α] [β]	Inverse Gamma distribution. There can be any number of value variables, all associated with the same parameter values. Parameters α and β can be variables, or if both are constant they can be specified in the constructor.

LessThan	out(B)
	in1(B,D,R)
	in2(B,D,R)
		-	Deterministic greater-than function, where out = in1 $<$ in2.

Name	Variables	Constructor	Description

LinearEquation	out(D,R)
	in(B,D,R)
		weights
		[smoothing]
			Deterministic linear equation, multiplying an input vector by a constant weight vector to equal the output variable. The weight vector is specified in the constructor. The number of *in* variables must equal the length of the weight vector. An optional smoothing value may be specified as a constructor argument[41].

Log	out(D,R)
	in(D,R)
		[smoothing]	Deterministic natural log function, where out = log(in). An optional smoothing value may be specified as a constructor argument[41].

LogNormal	[μ](R) [τ](R) value...(R)	[μ] [τ]	Log-normal distribution. There can be any number of value variables, all associated with the same parameter values. Parameters μ (mean) and $\tau = \frac{1}{\sigma^2}$ (precision) can be variables, or if both are constant then fixed parameters can be specified in the constructor.

MatrixProduct	C(D,R)
	A(D,R)
	B(D,R)
		N_r
		N_x
		N_c
		[smoothing]
			Deterministic matrix product function, $C = AB$, where A, B, and C are matrices. Constructor arguments, N_r specifies the number of rows in A and C, N_x specifies the number of columns in A and number of rows in B, and N_c specifies the number of columns in B and C. An optional smoothing value may be specified as a constructor argument[41].

MatrixVectorProduct	y(D,R)
	M(D,R)
	x(D,R)
		N_x
		N_y
		[smoothing]
			Deterministic matrix-vector product function, $y = Mx$, where x and y are vectors and M is a matrix. Constructor arguments, N_x and N_y, specify the input and output vector lengths, respectively. The matrix dimension is $N_y \times N_x$. An optional smoothing value may be specified as a constructor argument[41].

MatrixRealJoint
VectorProduct
	y(RJ)
	M(D,R)
	x(RJ)
		N_x
		N_y
		[smoothing]
			Deterministic matrix-vector product function, $y = Mx$, where x and y are RealJoint values and M is a matrix. Constructor arguments, N_x and N_y, specify the input and output vector lengths, respectively. The matrix dimension is $N_y \times N_x$. An optional smoothing value may be specified as a constructor argument[41].

Name	Variables	Constructor	Description
Multinomial	$[N]$(D) α(RJ) x...(D)	$[N]$	Multinomial distribution, $p(x\|N,\alpha)$, where N is the total number of trials, α is a vector of parameter variables, and x is a count of outcomes in each category. Parameter N can be a Discrete variable with positive integer values or a constant integer value specified in the constructor. The number of elements in α must exactly match the number of elements of x. The domain of each x variable must include integers from 0 through N, or if N is a variable, through the maximum value in the domain of N. The α parameters are represented as a normalized probability vector. The conjugate prior for this representation is a Dirichlet distribution[42].
Multinomial EnergyParameters	$[N]$(D) α...(R) x...(D)	D, $[N]$	Multinomial distribution, $p(x\|N,\alpha)$, where N is the total number of trials, α is a vector of parameter variables, and x is a count of outcomes in each category. Parameter N can be a Discrete variable with positive integer values or a constant integer value specified in the constructor. The number of elements in α must exactly match the number of elements of x, which must match the value of the constructor argument, D. The domain of each x variable must include integers from 0 through N, or if N is a variable, through the maximum value in the domain of N. In this alternative version of the Multinomial distribution, the α parameters are represented as energy values, that is, $\alpha = -\log(\rho)$, where ρ are unnormalized probabilities. The conjugate prior for this representation is such that each entry of α is independently distributed according to a negative exp-Gamma distribution, all with a common β parameter[42].

Name	Variables	Constructor	Description
Multinomial Unnormalized Parameters	$[N]$(D) α...(R) x...(D)	D, $[N]$	Multinomial distribution, $p(x\|N, \alpha)$, where N is the total number of trials, α is a vector of parameter variables, and x is a count of outcomes in each category. Parameter N can be a Discrete variable with positive integer values or a constant integer value specified in the constructor. The number of elements in α must exactly match the number of elements of x, which must match the value of the constructor argument, D. The domain of each x variable must include integers from 0 through N, or if N is a variable, through the maximum value in the domain of N. In this alternative version of the Multinomial distribution, the α parameters are represented as a vector of unnormalized probability values. The conjugate prior for this representation is such that each entry of α is independently distributed according to a Gamma distribution, all with a common β parameter[42].

Multiplexer	out(any)
	select
	in...(any)
		[smoothing]	Deterministic multiplexer[44]. The selector must be a discrete variable that selects one of the inputs to pass to the output. The data type of all inputs must be identical to that of the output. For RealJoint variables, the dimension of all inputs must equal that of the output. The with domain of the selector variable must be zero-based contiguous integers, $0...N-1$, where N is the number of input variables. An optional smoothing value may be specified as a constructor argument[41].

[44] Note that for the SumProduct solver, an optimized custom implementation of this factor function is used automatically, which avoids creation of a corresponding factor table.

Name	Variables	Constructor	Description
MultivariateNormal	value...(RJ)	μ Σ	Multivariate Normal distribution. There can be any number of value variables, all associated with the same parameter values. Parameters μ (mean vector) and Σ (covariance matrix) are constant that must be specified in the constructor[45]. The dimension of the mean vector, both dimensions of the covariance matrix, and the dimension of each value variable must be identical.

Negate	out(D,R)
	in(D,R)
		[smoothing]	Deterministic negation function, where out = -in. An optional smoothing value may be specified as a constructor argument[41].

NegativeExpGamma	$[\alpha]$(R) $[\beta]$(R) value...(R)	$[\alpha]$ $[\beta]$	Negative exp-Gamma distribution, which is a distribution over a variable whose negative exponential is Gamma distributed. That is, this is the negative log of a Gamma distributed variable. There can be any number of value variables, all associated with the same parameter values. Parameters α and β can be variables, or if both are constant they can be specified in the constructor, and correspond to the parameters of the underlying Gamma distribution.
Normal	$[\mu]$(R) $[\tau]$(R) value...(R)	$[\mu]$ $[\tau]$	Normal distribution. There can be any number of value variables, all associated with the same parameter values. Parameters μ (mean) and $\tau = \frac{1}{\sigma^2}$ (precision) can be variables, or if both are constant then fixed parameters can be specified in the constructor.

Not	out(B)
	in(B)
		-	Deterministic logical NOT of function, where out = in.

NotEquals	out(B)
	in...(B,D,R)
		-	Deterministic not-equals function, where out = ∼(in(1) == in(2) == ...).

Or	out(B)
	in...(B)
		-	Deterministic logical OR function, where out = OR(in...).

[45] In this version of Dimple, there is no support for variable parameters in the MultivariateNormal distribution.

Name	Variables	Constructor	Description
Poisson	$[\lambda]$(R) k(D)	$[\lambda]$	Poisson distribution, $p(k\|\lambda)$, where λ is the rate parameter, and k is the discrete output. While the value of k for a Poission distribution is unbounded, the domain should be set to include integers from 0 through a maximum value. The maximum value should be a multiple of the maximum likely value of λ[46]. The conjugate prior for the parameter, λ, is a Gamma distribution[42].

Power	out(D,R)
	base(D,R)
	power(D,R)
		[smoothing]	Deterministic power function, where out = base $^{\text{power}}$. An optional smoothing value may be specified as a constructor argument[41].

Product	out(D,R)
	in...(B,D,R)
		[smoothing]	Deterministic product function, where out = $\prod$ in. An optional smoothing value may be specified as a constructor argument[41].

Rayleigh	$[\sigma]$(R) value...(R)	$[\sigma]$	Rayleigh distribution. There can be any number of value variables, all associated with the same parameter value. Parameter σ can be a variable, or if constant, can be specified in the constructor.

RealAndImaginary ToComplex	out(C)
	inReal(R)
	inImag(R)
		[smoothing]	Deterministic conversion of two Real variables to a Complex variable, where the first input represents the real component and the second represents the imaginary component. An optional smoothing value may be specified as a constructor argument[41].

RealJointNegate	out(RJ)
	in(RJ)
		[smoothing]	Deterministic negation function for RealJoint variables, where out = -in. An optional smoothing value may be specified as a constructor argument[41].

RealJointProjection	out(R)
	in(RJ)
		index
		[smoothing]
			Deterministic conversion of a RealJoint variable to a Real variable corresponding to one specific element of the RealJoint variable. The *index* constructor argument indicates which element of the RealJoint variable to be used (using zero-based numbering). An optional smoothing value may be specified as a constructor argument[41].

[46] If the maximum value is 5 times larger than the largest value of λ, then less than 0.1 of the probability mass would fall above this value.

Name	Variables	Constructor	Description

RealJointSubtract	out(RJ)
	posIn(RJ)
	negIn...(RJ)
		[smoothing]	Deterministic subtraction function for RealJoint variables, where out = posIn $-$ $\sum$negIn. An optional smoothing value may be specified as a constructor argument[41].

RealJointSum	out(RJ)
	in...(RJ)
		[smoothing]	Deterministic summation function for RealJoint variables, where out = $\sum$in. An optional smoothing value may be specified as a constructor argument[41].

RealJointTo
RealVector
	out...(R)
	in(RJ)
		[smoothing]	Deterministic conversion of a RealJoint variable to a vector of Real variables. An optional smoothing value may be specified as a constructor argument[41].

RealVectorTo
RealJoint
	out(RJ)
	in...(R)
		[smoothing]	Deterministic conversion of a vector of Real variables to a RealJoint variable. An optional smoothing value may be specified as a constructor argument[41].

Sin	out(D,R)
	in(D,R)
		[smoothing]	Deterministic sine function, where out = sin(in). An optional smoothing value may be specified as a constructor argument[41].

Sinh	out(D,R)
	in(D,R)
		[smoothing]	Deterministic hyperbolic-sine function, where out = sinh(in). An optional smoothing value may be specified as a constructor argument[41].

Sqrt	out(D,R)
	in(D,R)
		[smoothing]	Deterministic square root function, where out = sqrt(in). An optional smoothing value may be specified as a constructor argument[41].

Square	out(D,R)
	in(D,R)
		[smoothing]	Deterministic square function, where out = in^2. An optional smoothing value may be specified as a constructor argument[41].

Subtract	out(D,R)
	posIn(B,D,R)
	negIn...(B,D,R)
		[smoothing]	Deterministic subtraction function, where out = posIn $-$ $\sum$negIn. An optional smoothing value may be specified as a constructor argument[41].

Sum	out(D,R)
	in...(B,D,R)
		[smoothing]	Deterministic summation function, where out = $\sum$in. An optional smoothing value may be specified as a constructor argument[41].

Tan	out(D,R)
	in(D,R)
		[smoothing]	Deterministic tangent function, where out = tan(in). An optional smoothing value may be specified as a constructor argument[41].

Tanh	out(D,R)
	in(D,R)
		[smoothing]	Deterministic hyperbolic-tangent function, where out = tanh(in). An optional smoothing value may be specified as a constructor argument[41].

Name	Variables	Constructor	Description

VectorInnerProduct	z(D,R)
	x(D,R,RJ)
	y(D,R,RJ)
		[smoothing]	Deterministic vector inner product function, $z = x\dot{y}$, where x and y are vectors and z is a scalar. Each vector input may be either an array of scalar variables, or a single RealJoint variable. The number of elements in x and y must be identical. An optional smoothing value may be specified as a constructor argument[41].

VonMises	$[\mu]$(R) $[\tau]$(R) value...(R)	$[\mu]$ $[\tau]$	Von Mises distribution. There can be any number of value variables, all associated with the same parameter values. Parameters μ (mean) and $\tau = \frac{1}{\sigma^2}$ (precision) can be variables, or if both are constant then fixed parameters can be specified in the constructor. The distribution is non-zero for value variables in the range $-\pi$ to π.

Xor	out(B)
	in...(B)
		-	Deterministic logical XOR function, where out = XOR(in...).

5.9.1 Factor Creation Utility Functions

Dimple includes some helper functions to create other built-in factors using a similar syntax to the overloaded MATLAB functions described in section 5.10. As for other overloaded functions, above, Dimple automatically creates the factors as well as the output variable(s). Such helper functions are defined for the following built-in distributions:

- Bernoulli
- Beta
- Binomial
- Categorical[47]
- CategoricalEnergyParameters
- Dirichlet
- ExchangeableDirichlet
- Gamma
- InverseGamma
- LogNormal

[47] This utility can be used to create either a Categorical or CategoricalUnnormalizedParameters factor, depending on whether the first argument is a RealJoint variable or an array of Real variables.

- Multinomial[48]
- MultinomialEnergyParameters
- MultivariateNormal[49]
- NegativeExpGamma
- Normal
- Poisson[50]
- Rayleigh
- VonMises

For each, the arguments are the parameters of the distribution. For example:

```
W = Gamma(alpha, beta);
X = Normal(mean, precision);
Y = Categorical(alpha);
Z = Rayleigh(sigma);
D = Dirichlet(alpha);
E = ExchangeableDirichlet(dimension, alpha);
```

The parameters can be variables, constants[51], or some of each[52].

By default, calling one of these functions creates a single output variable, and the factor is added to the most-recently created graph. But, optional arguments allow you to specify the dimensions of the array of output variables, or to specify the factor graph. These arguments can be in either order after the parameters. For example:

```
W = Gamma(alpha, beta, altGraph);
X = Normal(mean, precision, [100, 1]);
Y = Categorical(alpha, [10, 10, 2], aGraph);
Z = Rayleigh(sigma, myGraph, size(somethingElse));
```

Dimple also includes built-in helper functions to create other factors that aren't distributions. Specifically:

[48] This utility can be used to create either a Multinomial or MultinomialUnnormalizedParameters factor, depending on whether the first argument is a RealJoint variable or an array of Real variables.

[49] In the current version of Dimple, only constant parameters are supported by the MultivariateNormal factor.

[50] This utility takes an argument λ, either a variable or constant, plus an additional argument $maxK$, which indicates the maximum likely value of the output variable, k. The value of $maxK$ should be a multiple of the maximum likely value of λ (if $maxK$ is 5 times larger than the largest value of λ, then less than 0.1 of the probability mass would be greater than $maxK$).

[51] If *all* parameter inputs are constants, then instead of creating both variables and factors, the variables are created, and the Input to each variable is set accordingly.

[52] For cases where a parameter is a vector of real values, the parameter may be a Real variable array, a cell-arrays that may mix Real variables and constants, or an array of constants.

```
outValue = If(condition, trueValue, falseValue);
```

The “If” helper function implements a conditional assignment function using the built-in Multiplexer factor. The function creates an output variable (or array of variables) that is of the same type and dimension as the trueValue and falseValue variables. These two input variables must either be of identical type, domain, and dimensions, or one of these may be a constant. The condition must be a Bit variable or array of variables. If trueValue and falseValue are arrays, the function creates an array of Multiplexer factors, each connected to the corresponding inputs and outputs. In the case of an array, if condition is a scalar Bit variable, all of these are connected to the same condition bit. Otherwise, condition must be of the same dimension as the other inputs. As with the other factor creation utilities, the function optionally takes a graph input to specify a particular graph in which to add the factors.

Dimple also includes some built-in helper functions to create structured graphs, combining smaller factors to form an efficient implementation of a larger subgraph. Specifically, the following functions are provided:

- getNBitXorDef(n), where n is a positive integer. Returns a nestable graph and an array of n-Bit connector variables. Efficient tree implementation of the XORDelta function.
- getVXOR(n), where n is a positive integer. Returns a nestable graph and an array of n-Bits connector variables. Constrains exactly one bit to be 1, and all others to be 0.
- MultiplexerCPD(domains) - See section 4.6.1
- MultiplexerCPD(domain, numZs) - See section 4.6.1

5.10 List of Overloaded MATLAB Operators and Functions

The following table lists the set of overloaded MATLAB operators that can be used to implicitly create factors. The table shows the operator, the corresponding built-in factor (as described in section 5.9), the valid variable data types of the inputs and outputs (B = Bit, D = Discrete, or R = Real, C = Complex, RJ = RealJoint), and wether or not vectorized inputs are supported. The use of these operators and functions is described in section 4.1.4.7.

Operator	Factor	Out	In1	In2	Vectorized	Description
&	And	B	B	B	✓	Logical AND
|	Or	B	B	B	✓	Logical OR
xor()	Xor	B	B	B	✓	Logical XOR
∼	Not	B	B	-	✓	Logical NOT
+	Sum	D,R[53]	D,R	D,R	✓	Plus
	ComplexSum	C	C,R	C,R	✓	Complex plus
	RealJointSum	RJ	RJ	RJ	-	RealJoint plus
−	Subtract	D,R[53]	D,R	D,R	✓	Minus
	ComplexSubtract	C	C,R	C,R	✓	Complex minus
	RealJointSubtract	RJ	RJ	RJ	-	RealJoint minus
	Negate	D,R[53]	D,R	-	✓	Unary minus
	ComplexNegate	C	C, R	-	✓	Unary complex minus
	RealJointNegate	RJ	RJ	-	-	Unary RealJoint minus
∗	Product	D,R[53]	D,R	D,R	✓[54]	Scalar multiply
	ComplexProduct	C	C,R	C,R	✓[54]	Complex scalar multiply
	VectorInnerProduct	R	D,R,RJ	D,R,RJ	-	Vector inner product[55]
	MatrixProduct	R	D,R	D,R	-	Matrix multiply[56]
	MatrixVectorProduct	R	D,R	D,R	-	Matrix-vector multiply[57]
	MatrixRealJoint VectorProduct	RJ	RJ,D,R	RJ,D,R	-	Matrix-vector multiply[58]
.∗	Product	D,R[53]	D,R	D,R	✓	Point-wise multiply
	ComplexProduct	C	C,R	C,R	✓	Complex pointwise multiply
/	Divide	D,R[53]	D,R	D,R	✓[59]	Scalar divide
	ComplexDivide	C	C,R	C,R	✓[59]	Complex scalar divide
./	Divide	D,R[53]	D,R	D,R	✓	Point-wise divide
	ComplexDivide	C	C,R	C,R	✓	Complex pointwise divide
∧	Power	D,R[53]	D,R	D,R	✓[60]	Scalar power
	Square[61]	D,R[53]	D,R	D,R	✓[60]	Scalar square

[53] If either input is Real, then the output is Real

[54] One of the inputs may be a vector as long as the other is a scalar.

[55] If both inputs are vectors of the same dimension, then the VectorInnerProduct factor will be used. Each vector may be an array of scalar variables or a RealJoint variable.

[56] If both inputs are two-dimensional matrices of appropriate dimension, then the MatrixProduct factor will be used.

[57] If one input is a vector and the other is a matrix of appropriate dimension, then the MatrixVectorProduct factor will be used.

[58] If one input is a RealJoint variable and the other is a matrix of scalar variables or constants of appropriate dimension, then the MatrixRealJointVectorProduct factor will be used.

[59] The dividend may be a vector as long as the divisor is a scalar.

[60] The base may be a vector as long as the exponent is a scalar.

[61] If the power is the constant 2, the Square factor is used instead of the Power factor.

Operator	Factor	Out	In1	In2	Vectorized	Description
.∧	Power	D,R[53]	D,R	D,R	✓	Point-wise power
	Square[61]	D,R[53]	D,R	D,R	✓	Point-wise square
′	ComplexConjugate	C	C	-	✓	Complex conjugate
<	LessThan	B	D,R	D,R	✓	Less than
>	GreaterThan	B	D,R	D,R	✓	Greater than
<=	GreaterThan[62]	B	D,R	D,R	✓	Less than or equal to
>=	LessThan[63]	B	D,R	D,R	✓	Greater than or equal to
Equals()	Equals	B	B,D,R	B,D,R[64]	✓	Equals[65]
NotEquals()	NotEquals	B	B,D,R	B,D,R[64]	✓	Not equals[66]
mod()	-	D	D	D	✓	Modulo function[67]
abs()	Abs	D,R	D,R	-	✓	Absolute value
	ComplexAbs	R	C	-	✓	Complex absolute value
sqrt()	Sqrt	R	R	-	✓	Square root
log()	Log	R	R	-	✓	Natural log
exp()	Exp	R	R	-	✓	Exponential function
	ComplexExp	C	C	-	✓	Complex exponential
sin()	Sin	R	R	-	✓	Sine
cos()	Cos	R	R	-	✓	Cosine
tan()	Tan	R	R	-	✓	Tangent
asin()	ASin	R	R	-	✓	Arc-sine
acos()	ACos	R	R	-	✓	Arc-cosine
atan()	ATan	R	R	-	✓	Arc-tangent
sinh()	Sinh	R	R	-	✓	Hyperbolic sine
cosh()	Cosh	R	R	-	✓	Hyperbolic cosine
tanh()	Tanh	R	R	-	✓	Hyperbolic tangent

[62] Uses GreaterThan factor, reversing the order.

[63] Uses LessThan factor, reversing the order.

[64] This function is not limited to two inputs, but can take an arbitrary number of inputs

[65] Equivalent to the == operator, but the == operator is not overloaded for this purpose so that it can instead be used to determine whether or not two variables reference the same Dimple variable.

[66] Equivalent to the ∼= operator, but the ∼= operator is not overloaded for this purpose so that it can instead be used to determine whether or not two variables reference the same Dimple variable.

[67] Currently, the mod() operator supports discrete variables only, and it uses the MATLAB definition of mod on negative numbers. This may be subject to change in future versions.

5.11 Other Top Level Functions

5.11.1 setSolver

```
setSolver('SolverName');
```

This function changes the default solver to the solver designated by the argument, which is a string indicating the name of the solver (see section 5.1.2.1 for the list of valid solver names, and section 5.6 for a description of each solver). The solver name is case insensitive.

5.11.2 dimpleVersion

Dimple provides a method to return a string describing the current Dimple version.

```
version = dimpleVersion();
```

This will produce something of the form:

```
<Release Number> <Git Branch Name> YYY-MM-DD HH:MM:SS <Timezone Offset>
```

For example:

```
0.04 master 2013-10-11 14:00:05 -0400
```

The date in the version string represents the last git commit date associated with the compiled code.

5.11.3 dimpleOptions

Returns a complete list option keys, corresponding to all of available Dimple options. See section 5.4 for a description of the Dimple options mechanism.

```
dimpleOptions()
```

A A Short Introduction to Factor Graphs

We introduce factor graphs, a powerful tool for statistical modeling. Factor graphs can be used to describe a number of commonly used statistical models, such as hidden Markov models, Markov random fields, Kalman filters, and Bayes nets.

Suppose that we are given a set of n discrete random variables: $a_1, ..., a_n$. The random variables have some joint probability distribution: $p(a_1, a_2, ..., a_n)$. Suppose that the joint probability distribution factors, in the following sense: there exist subsets $S_1, ..., S_k \subseteq \{1, 2, ..., n\}$ where $S_j = \{S^j_1, s^j_2, ..., s^j_{t(j)}\}$ and such that $p(a_1, a_2, ..., a_n) = \prod_{j=1}^{k} f_j(a_{s^j_1}, a_{s^j_2}, ..., a_{s^j_{t(j)}})$.

For example, if the a_i form a Markov chain, then the joint probability can be factored as

$$\begin{aligned} p(a_1, a_2, ..., a_n) &= p(a_1) \prod_{j=1}^{n-1} p(a_{j+1}|a_j) && (1) \\ &= f_0(a_1) \prod_{j=1}^{n-1} f_j(a_j, a_{j+1}) && (2) \end{aligned}$$

The factors above are normalized, in the sense that as the a_i vary, the probabilities sum to one. We will define our factors more generally and ask only that they are proportional to the joint probability. So, we call the f_j a collection of factors of p if

$$p(a_1, a_2, ..., a_n) \propto \prod_{j=1}^{k} f_j(a_{s^j_1}, a_{s^j_2}, ..., a_{s^j_{t(j)}})$$

The product of the factors then differs from the joint probability only by multiplication by a normalizing constant.

When a probability distribution can be expressed as a product of small factors (i.e., $|S_j|$ is small for all j), then if is possible to invoke a host of powerful tools for modeling and inference, as we will soon see.

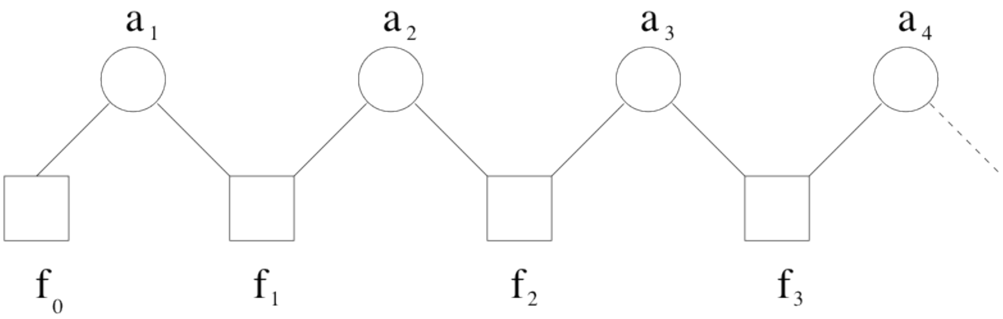
Suppose that we are given a factored representation of a joint probability distribution. It is possible to describe the structure of the factors as a graph. We can represent each variable a_i and each function f_j by a node in the graph, and place an (undirected) edge between node a_i and node f_j if and only if the variable a_i is an argument in the function f_j. These two types of nodes are referred to as factor nodes and variable nodes. Because all edges lie between the two disjoint classes of nodes, the resulting graph is bipartite. This graph is called a factor graph.

In the remainder of this documentation, we slightly abuse notation and use f_j and a_i to refer both to the nodes of the factor graph and to the underlying factors and variables (i.e., both the graphical representation of these entities and the mathematical entities underlying them).

To understand what factor graphs look like, we will construct several examples. First, let us continue with a Markov chain. Equation 1 expressed a Markov chain in factored form, where

$$f_j(a_j, a_{j+1}) = p(a_{j+1}|a_j)$$

We display the corresponding factor graph in the following figure:



Next, let us consider a hidden Markov model (HMM). We can construct the corresponding factor graph by extending the previous example. An HMM contains a Markov chain transiting from state a_i to a_{i+1}. There is also an observation b_i made of each state; if we are given a_i, then b_i is conditionally independent of all other variables. We can incorporate this probability by using a factor:

$$g_i(a_i) = Pr(b_i|a_i)$$

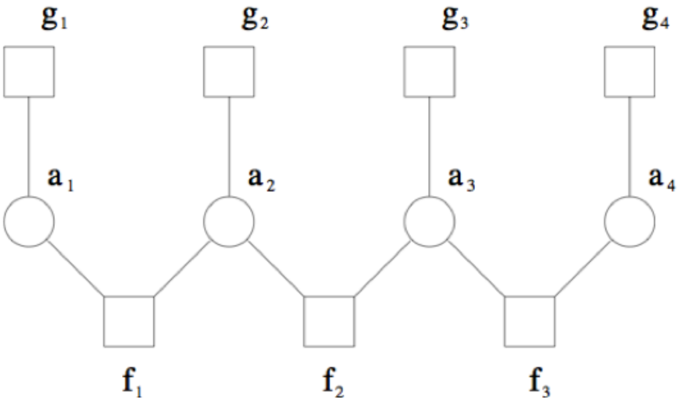
The product of our factors is then

$$\begin{aligned} f_0 \left(\prod_{j=1}^{n-1} f_j(a_j, a_j + 1) \right) \prod_{j=1}^{n} g_j(a_j) &= Pr(a_1) \left(\prod_{j=1}^{n-1} Pr(a_{j+1}|a_j) \right) \prod_{j=1}^{n} Pr(b_j|a_j) \\ &= Pr(a_1, ..., a_n, b_1, ..., b_n) \end{aligned}$$

Since the b_i observed, then $Pr(b_1, ..., b_n)$ is a constant. Therefore

$$\begin{aligned} Pr(a_1, ..., a_n, b_1, ..., b_n) &\propto \frac{Pr(a_1, ..., a_n, b_1, ..., b_n)}{Pr(b_1, ..., b_n)} \\ &= Pr(a_1, ..., a_n|b_1, ..., b_n) \end{aligned}$$

as desired.



The resulting factor graph takes the following form illustrated in the figure above. Note that the variables b_i need not appear explicitly in the factor graph; we have incorporated their effect in the g_i factors.

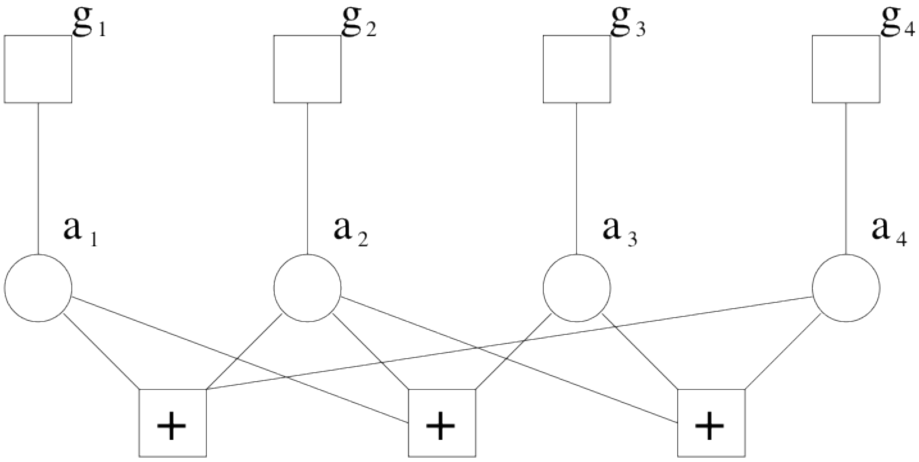
Generalizing from a Markov chain to an HMM illustrates a very powerful feature of factor graphs. Complicated mathematical models are often composed of simpler parts. When these models are expressed as factor graphs, we can frequently reuse the simpler factor graphs to construct the more complicated ones. This can be done simply in Dimple by using the nested graphs feature (see section 3.3 on Nested Graphs).

As a final example, we will construct a factor graph for error correction (for this more advanced topic, we will assume the reader is familiar with LDPC codes). Suppose that we receive a codeword from a 4-bit LDPC error-correcting code that has been corrupted by noise. The sender wishes to communicate a four-bit codeword (a_1, a_2, a_3, a_4) satisfying some parity check equations, but the receiver only observes the corrupted values (b_1, b_2, b_3, b_4). (The domain of the b_i is determined by the communication channel. For instance, if we have a discrete binary symmetric channel, then the b_i will be bits; if we have a continuous additive white Gaussian noise channel and some modulation scheme, the b_i will be real-valued.) Let H be the parity check matrix of the LDPC code used, i.e., the codeword (a_1, a_2, a_3, a_4) verifies the equation $Ha = 0 \bmod 2$.

For instance, suppose that H is the following parity check matrix:

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

Suppose that the error is independent, i.e., if we condition on a_i, b_i is conditionally independent of the other variables. Then, the following factor graph represents the decoding task at hand.



The construction above applies to any linear binary ECC. However, if every row and column of H is sparse (as would be the case with an LDPC code), then every factor is small, and every node in the factor graph will be of small degree.

Given a factor graph, the objective is often to compute the marginal distribution of the random variables a_i of the graph (this also allows us to find the most likely value taken by each variable, by maximization of the marginal probability). Dimple provides an implementation of belief propagation (BP) (in its sum-product version), in order to approximately compute the marginal distribution of each random variable.

BP is an iterative message-passing algorithm where messages pass along the edges of the factor graph. A "message" can be viewed as an un-normalized probability distribution. The algorithm comes in a number of variants depending on the message update rule and the order of the message updates.

The sum-product form of BP generalizes a number of algorithms, including the "forward-backward" algorithm of HMMs, and the BCJR algorithm of coding theory. It always gives the exact answer when the underlying factor graph is a tree (if the graph contains no cycles). Although it is not an exact algorithm for general graphs, BP has been found to give excellent results for a wide variety of factor graphs, and runs particularly fast on sparse factor graphs (i.e., factor graphs of low node degree).

B Creating Custom Dimple Extensions in Java

B.1 Creating a Custom Factor Function

There are some cases in which it is desirable to add a custom factor function written in Java rather than MATLAB. Specific cases where this is desirable are:

- A factor function is needed to support continuous variables that is not available as a Dimple built-in factor.
- A MATLAB factor function runs too slowly when creating a factor table, where a Java implementation may run more quickly.

To create a custom factor function, you must create a Java class that extends the Dimple `FactorFunction` class. When extending the `FactorFunction` class, the following method must be overwritten:

- `evalEnergy`: Evaluates a set of input values and returns an energy value (negative log of a weight value).

The user may extend other methods, as appropriate:

- Constructor: If a constructor is specified (for example, to pass constructor arguments), it must call the constructor of the super class.
- `isDirected`: Indicates whether the factor function is directed. If directed, then there are a set of directed outputs for which the marginal distribution for all possible input values is a constant. If not overridden, this is assumed false.
- `getDirectedToIndices`: If a factor function is directed, indicates which edge indices are the directed outputs (numbering from zero), returning an array of integers. There are two forms of this method, which may be used depending on whether the set of directed outputs depends on the number of edges in the factor that uses this factor function (many factor functions support a variable number of edges). If `isDirected` is overridden and can return `true`, then this method must also be overridden.
- `isDeterministicDirected`: Indicates whether a factor function is both directed and deterministic. If deterministic and directed, then it is in the form of a deterministic function such that for all possible settings of the input values there is exactly one output value the results in a non-zero weight (or, equivalently, a non-infinite energy)[68]. If not overridden, this is assumed false.
- `evalDeterministic`: If a factor function is directed and deterministic, this method evaluates the values considered the inputs of the deterministic function and returns

[68] The indication that a factor function is deterministic directed is used by the Gibbs solver, and is necessary for such factor functions to work when using the Gibbs solver.

the resulting values for the corresponding outputs. Note that these are not the weight or energy values, but the actual values of the associated variables that are considered outputs of the deterministic function. If `isDeterministicDirected` is overridden and can return `true`, then this method must also be overridden.

- `eval`: Evaluates a set of input values and returns a weight instead of an energy value. Overriding this method would only be useful if implementing this method can be done significantly more computationally efficiently than the default implementation, which calls evalEnergy and then computes $\exp(-energy)$.

The following is a very simple example of a custom factor function:

```
import com.analog.lyric.dimple.factorfunctions.core.FactorFunction;

/*
 * This factor enforces equality between all variables and weights
 * elements of the domain proportional to their value
 */
public class BigEquals extends FactorFunction
{
    @Override
    public final double evalEnergy(Value[] input)
    {
        if (input.length == 0)
            return 0;

        Value firstVal = input[0];

        for (int i = 1; i < input.length; i++)
            if (!input[i].valueEquals(firstVal))
                return Double.POSITIVE_INFINITY;

        return 0;
    }
}
```

B.2 Creating a Custom Proposal Kernel

In some cases, it may be useful to add a custom proposal kernel when using the Gibbs solver with a Metropolis-Hastings sampler. In particular, since the *block* Metropolis-Hastings sampler does not have a default proposal kernel, it is necessary to add a custom proposal kernel in this case.

To create a custom proposal kernel, you must create a Java class that implements either the Dimple `IProposalKernel` interface in the case of a single-variable proposal kernel, or the `IBlockProposalKernel` interface in the case of a block proposal kernel.

These interfaces define the following methods that must be implemented:

- `next` This method takes the current value(s) of the associated variable(s) along with the corresponding variable domain(s), and returns a proposal. The proposal object returned (either a `Proposal` object for a single-variable proposal or a `BlockProposal`: object for a block proposal) includes both the proposed value(s) as well as the forward and reverse proposal probabilities (the negative log of these probabilities).
- `setParameters`: Allows the class to include user-specified parameters that can be set to modify the behavior of the kernel. This method must be implemented but may be empty if no parameters are needed.
- `getParameters`: Returns the value of any user-specified parameters that have been specified. This method must be implemented but may return null if no parameters are needed.

B.3 Compiling Dimple Extensions in Java

The new class must be compiled to class files, or optionally create a jar file. If using Eclipse, users can simply create a new project, create the new class, and the .class files will be created automatically.

B.4 Adding Java Binary to MATLAB Path

In MATLAB, you must use the javaaddpath call to add the Java class or jar files to the javaclasspath. For example:

```
javaaddpath('<path to my project>/MyFactorFunctions/bin');
```

or

```
javaaddpath('<path to the jar>/myjar.jar');
```