Reconstructing the MSSM Lagrangian from LHC data

Michael Rauch
Rémi Lafaye, Tilman Plehn, Dirk Zerwas
What SFitter does

Set of measurements

- LHC measurements:
  - kinematic edges, thresholds, masses, mass differences
  - cross sections, branching ratios
- ILC measurements
- Indirect Constraints
  - electro-weak: $M_W$, $\sin^2 \theta_W$; \hspace{1cm} (g − 2)$_\mu$
  - flavour: BR($b \rightarrow s \gamma$), BR($B_s \rightarrow \mu^+ \mu^-$); \hspace{1cm} dark matter: $\Omega h^2$

or even ATLAS and CMS measurements separately

Compare to theoretical predictions

- Spectrum calculators: SoftSUSY, SuSPECT, ISASUSY
  \hspace{1cm} \[\text{Allanach; Djouadi, Kneur, Moultaka; Baer, Paige, Protopopescu, Tata}\]
- LHC cross sections: Prospino2 \[\text{Plehn et al.}\]
- LC cross sections: MsmLib \[\text{Ganis}\]
- Branching Ratios: SUSYHit (HDecay + SDecay) \[\text{Djouadi, Mühleitner, Spira}\]
- micrOMEGAs \[\text{Bélanger, Boudjema, Pukhov, Semenov}\]
- g-2 \[\text{Stöckinger}\]
Parameter Scans

- MSSM parameter space is high-dimensional:
  - SM: 3+ parameters \((m_t, \alpha_s, \alpha, \ldots)\)
  - mSUGRA: 5 parameters \((m_0, m_{1/2}, A_0, \tan(\beta), \text{sgn}(\mu))\)
  - General MSSM: 105 parameters

- On loop-level observables depend on every parameter
  Simple inversion of the relations not possible
  \Rightarrow Parameter scans

- Error estimates on parameters in the minimum

Find best points (best \(\chi^2\)) using different fitting techniques:

- fixed Grid scan
  \(+\) scans complete parameter space
  \(-\) many points needed \((O(e^N))\)

- Gradient search (Minuit)
  \(+\) Reasonably fast
  \(-\) Limited convergence, only best fit

- Weighted Markov Chains
Markov Chain (MC):

- Sequence of points, chosen by an algorithm (Metropolis-Hastings), only depending on its direct predecessor
- Picks a set of "average" points according to a potential $V$ (e.g. inverse log-likelihood, $1/\chi^2$)
- Point density resembles the value of $V$ (i.e. more points in region with high $V$)
- Scans high dimensional parameter spaces efficiently [Baltz, Gondolo 2004]
- mSUGRA MC scans with current exp. limits [Allanach, Lester, Weber 2005-7; Roszkowski, Ruiz de Austro, Trotta 2006/7]

Weighted Markov Chains: Improved evaluation algorithm for binning: [Plehn, MR]

- Weight points with value of $V$: 
  \[
  \left( \frac{\text{number of points}}{\sum_{\text{points}} 1/V(\text{point})} \right)
  \]
  [based on Ferrenberg, Swendsen 1988]
- Maintain additional chain which stores points rejected because $V(\text{point}) = 0$
  + Fast scans of high-dimensional spaces $O(N)$
  + Does not rely on shape of $\chi^2$ (no derivatives used)
  + Can find secondary distinct solutions
  - Exact minimum not found $\Rightarrow$ Additional gradient fit
  - Bad choice of proposal function for next point leads to bad coverage of the space
mSUGRA as a Toy Model

mSUGRA with LHC measurements (SPS1a kinematic edges):
pick one set of ”measurements”, randomly smeared from the true values

Free parameters:
\( m_0, m_{1/2}, \tan(\beta), A_0, \text{sgn}(\mu), m_t \)

SFitter output 1:
Fully-dimensional exclusive likelihood map
(colour: minimum \( \chi^2 \) over all unseen parameters)

SFitter output 2:
Ranked list of minima:

\[
\begin{array}{cccccccc}
\chi^2 & m_0 & m_{1/2} & \tan(\beta) & A_0 & \mu & m_t \\
1) & 1.32 & 100.4 & 251.2 & 12.7 & -71.7 & + & 171.9 \\
2) & 7.18 & 106.3 & 243.6 & 14.3 & -103.3 & – & 170.7 \\
3) & 13.9 & 103.5 & 258.2 & 12.2 & 848.4 & + & 174.4 \\
4) & 75.1 & 107.3 & 251.4 & 15.1 & 778.8 & – & 173.6 \\
\ldots
\end{array}
\]
Bayesian or Frequentist?

SFitter provides full-dimensional log-likelihood map
→ "project" onto plotable 1- or 2-dimensional spaces

Bayesian:

![Bayesian likelihood map](image1)

Marginalisation of $\chi^2$ in all other directions

Frequentist:

![Frequentist likelihood map](image2)

Profile likelihood: Value of bin is value of smallest $\chi^2$ occurring in this bin
Bayesian or Frequentist?

SFitter provides full-dimensional log-likelihood map
→ "project" onto plotable 1- or 2-dimensional spaces

Bayesian:

Frequentist:

Marginalisation of $\chi^2$ in all other directions

Profile likelihood: Value of bin is value of smallest $\chi^2$ occurring in this bin

Different methods answer different questions.
⇒ Bayesian and Frequentist!
Everybody can choose his/her favourite analysis . . .
Purely high-scale model

\[ m_0, m_{1/2}, A_0 \text{ defined at the GUT-scale } \Leftrightarrow \tan(\beta) \text{ defined at the weak scale} \]
\[ \Rightarrow \text{Replace } \tan(\beta) \text{ with high-scale quantity } B \]
\[ \Rightarrow \text{Flat prior in } B \text{ yields prior } \propto \frac{1}{\tan(\beta)^2} \]

SPS1\(\alpha\) with LHC kinematic edges (\(\tan(\beta)\) vs. \(1/\chi^2\)):

**Flat prior**

Bayesian:
Large influence of choice of prior
Choosing flat \(B\) prior strongly favours low values of \(\tan(\beta)\).

Frequentist:
Two plots should be identical
(no prior in \(\chi^2\) calculation)
Indirect influence via Markov Chain proposal function
Error determination

Treatment of errors:

- All experimental errors are Gaussian
  \[ \sigma_{\text{exp}}^2 = \sigma_{\text{stat}}^2 + \sigma_{\text{syst}(j)}^2 + \sigma_{\text{syst}(l)}^2 \]

- Systematic errors from jet (\(\sigma_{\text{syst}(j)}\)) and lepton energy scale (\(\sigma_{\text{syst}(l)}\)) assumed 99% correlated each

- Theory error added as box-shaped (RFit scheme \([\text{Hoecker, Lacker, Laplace, Lediberder}]\))
  \[ \Rightarrow -2 \log L \equiv \chi^2 = \sum_{\text{measurements}} \left\{ \begin{array}{ll} 0 & \text{for } |x_{\text{data}} - x_{\text{pred}}| < \sigma_{\text{theo}} \\ \left( \frac{|x_{\text{data}} - x_{\text{pred}}| - \sigma_{\text{theo}}}{\sigma_{\text{exp}}} \right)^2 & \text{for } |x_{\text{data}} - x_{\text{pred}}| \geq \sigma_{\text{theo}} \end{array} \right. \]

⇒ Parameter errors:

<table>
<thead>
<tr>
<th>SPS1a</th>
<th>(\Delta_{\text{theo-expect}}^\text{flat})</th>
<th>(\Delta_{\text{theo-expect}}^\text{zero})</th>
<th>(\Delta_{\text{theo-expect}}^\text{gauss})</th>
<th>(\Delta_{\text{theo-expect}}^\text{flat})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LHC masses</td>
<td>LHC edges</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m_0)</td>
<td>100</td>
<td>4.89</td>
<td>0.50</td>
<td>2.96</td>
</tr>
<tr>
<td>(m_{1/2})</td>
<td>250</td>
<td>3.27</td>
<td>0.73</td>
<td>2.99</td>
</tr>
<tr>
<td>(\tan \beta)</td>
<td>10</td>
<td>2.73</td>
<td>0.65</td>
<td>3.36</td>
</tr>
<tr>
<td>(A_0)</td>
<td>-100</td>
<td>56.4</td>
<td>21.2</td>
<td>51.5</td>
</tr>
<tr>
<td>(m_t)</td>
<td>171.4</td>
<td>0.98</td>
<td>0.26</td>
<td>0.89</td>
</tr>
</tbody>
</table>

⇒ Use kinematic edges for parameter determination instead of masses
Weak-scale MSSM

- No need to assume specific SUSY-breaking scenario
- Use of Markov Chains makes scanning the 19-dimensional parameter space feasible
- Lack of sensitivity on one parameter does not slow down the scan (no need to fix parameters)
- Underdetermined combinations of parameters can wash out correlations in plots
- Same SFitter output as before: Minima list and Likelihood map

MSSM using SPS1a spectrum and LHC kinematic edges:
($M_1, M_2, M_3, \mu, \tan(\beta), m_t$ sub-space)

Marginalisation  Profile likelihood
Weak-scale MSSM

MSSM using SPS1a spectrum and LHC kinematic edges:
(upper line: $M_1$, $M_2$, $M_3$, $\mu$, $\tan(\beta)$, $m_t$ sub-space
lower line: full parameter space)

Marginalisation

Profile likelihood
Dark Matter

Content of the universe:
- 73% Dark energy
- 4% Ordinary matter
- 23% Dark matter

MSSM: $\chi_1^0$ as LSP ideal candidate for cold dark matter (CDM): massive, weakly interacting

- SFitter: Determine Lagrangian parameters $\Rightarrow$ Spectrum and couplings
- e.g. micrOMEGAs: Calculate relic density $\Omega_{\text{CDM}} h^2 = n_{\text{LSP}} m_{\text{LSP}}$

$\Rightarrow$ Prediction of $\Omega_{\text{CDM}} h^2$

LHC: $\Omega_{\text{CDM}} h^2 = 0.1906 \pm 0.0033$
LHC+ILC: $\Omega_{\text{CDM}} h^2 = 0.1910 \pm 0.0003$
(improvement by one order of magnitude)

Compare with experiment
(Measurement of the fluctuations of the cosmic microwave background):

WMAP: $\Omega_{\text{CDM}} h^2 = 0.1277 \pm 0.008$ [astro-ph/0603449]
Planck: $\Omega_{\text{CDM}} h^2 = ? \pm 0.0016$
Parameter scans important to determine Lagrangian parameters from observables

Problem of high-dimensional parameter spaces

Markov Chains can do this effectively

Improved algorithm developed

Two types of output: Likelihood map and list of best points

Both Bayesian and Frequentist from likelihood map

Bayesian output significantly dependent on priors

Tested with mSUGRA SPS1a:
- can reconstruct SPS1a from (simulated) LHC data

Repeated procedure with weak-scale MSSM:
- reconstruction works as well

SFitter (despite its name) not tied to SUSY
- extend to other models/problems
Backup Slides
Metropolis-Hastings Algorithm

1. Start point
2. Probability density function (PDF)
3. Suggested point
4. \[ \frac{V(\text{suggested point})}{V(\text{start point})} \] ? random number \( \in [0, 1) \)
   - Yes: replace
   - No: add point to chain
mSUGRA around Minima

\[ \text{sgn}(\mu) = +1 \]

\[ \text{sgn}(\mu) = -1 \]

\[ A_0 \]

\[ m_t \]
mSUGRA: Correct solution vs. negative $\mu$ solution

Experimental results smeared by random number distributed as Gaussian around central value

[plot by D Zerwas 2006]
mSUGRA SPS1a as a benchmark point:
\[ m_0 = 100 \text{ GeV}, \quad m_{1/2} = 250 \text{ GeV}, \quad \tan \beta = 10, \quad A_0 = -100 \text{ GeV}, \quad \mu > 0, \text{ and } \quad m_{\text{top}} = 174.1 \text{ GeV} \]

The LHC “experimental” data from cascade decays:

Theoretical errors:
- 3% for gluino and squark masses
- 1% for other sparticle masses
### Experimental Input (edges)

<table>
<thead>
<tr>
<th>(Obs)</th>
<th>= (meas) ± (exp) ± (theo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{h^0}$</td>
<td>$109.53 \pm 0.25 \pm 2.0$</td>
</tr>
<tr>
<td>$m_t$</td>
<td>$171.4 \pm 1.0 \pm 0.0$</td>
</tr>
<tr>
<td>$\Delta m_{\tilde{\mu}_L, \chi_1^0}$</td>
<td>$106.26 \pm 1.6 \pm 0.1$</td>
</tr>
<tr>
<td>$\Delta m_{\tilde{g}, \chi_1^0}$</td>
<td>$509.96 \pm 2.3 \pm 6.0$</td>
</tr>
<tr>
<td>$\Delta m_{\tilde{c}_R, \chi_1^0}$</td>
<td>$450.52 \pm 10.0 \pm 4.2$</td>
</tr>
<tr>
<td>$\Delta m_{\tilde{g}, \tilde{b}_1}$</td>
<td>$98.971 \pm 1.5 \pm 1.0$</td>
</tr>
<tr>
<td>$\Delta m_{\tilde{g}, \tilde{b}_2}$</td>
<td>$64.016 \pm 2.5 \pm 0.7$</td>
</tr>
<tr>
<td>$\text{Edge}(\chi_2^0, \tilde{\mu}_R, \chi_1^0)$</td>
<td>$79.757 \pm 0.03 \pm 0.08$ ($m_{ll}^{\text{max}}$)</td>
</tr>
<tr>
<td>$\text{Edge}(\tilde{c}_L, \chi_2^0, \chi_1^0)$</td>
<td>$446.44 \pm 1.4 \pm 4.3$  ($m_{llq}^{\text{max}}$)</td>
</tr>
<tr>
<td>$\text{Edge}(\tilde{c}_L, \chi_2^0, \tilde{\mu}_R)$</td>
<td>$316.51 \pm 0.9 \pm 3.0$  ($m_{lq}^{\text{low}}$)</td>
</tr>
<tr>
<td>$\text{Edge}(\tilde{c}_L, \chi_2^0, \tilde{\mu}_R, \chi_1^0)$</td>
<td>$392.8 \pm 1.0 \pm 3.8$  ($m_{lq}^{\text{high}}$)</td>
</tr>
<tr>
<td>$\text{Edge}(\chi_4^0, \tilde{\mu}_R, \chi_1^0)$</td>
<td>$257.41 \pm 2.3 \pm 0.3$  ($m_{ll}^{\text{max}}(\chi_4^0)$)</td>
</tr>
<tr>
<td>$\text{Edge}(\chi_4^0, \tilde{\tau}_L, \chi_1^0)$</td>
<td>$82.993 \pm 5.0 \pm 0.8$  ($m_{\tau\tau}^{\text{max}}$)</td>
</tr>
<tr>
<td>$\text{Threshold}(\tilde{c}_L, \chi_2^0, \tilde{\mu}_R, \chi_1^0)$</td>
<td>$211.95 \pm 1.6 \pm 2.0$  ($m_{llq}^{\text{min}}$)</td>
</tr>
<tr>
<td>$\text{Threshold}(\tilde{b}_1, \chi_2^0, \tilde{\mu}_R, \chi_1^0)$</td>
<td>$211.95 \pm 1.6 \pm 2.0$  ($m_{llb}^{\text{min}}$)</td>
</tr>
</tbody>
</table>
Error determination

Minuit output not usable for flat theory errors:

- Migrad function depends on parabolic approximation
- Cannot determine $\Delta \chi^2$ for Minos to yield 68% CL intervals

⇒ Need more general approach

- Perform 10,000 toy experiments with measurements smeared around correct value
- Minimise each toy experiment
- Plot resulting distribution of parameter points and fit with Gaussian

Flat theory errors

Gaussian theory errors
Example

Test function (5-dim):

- **Small Hypersphere** $r = 100$, $V_{\text{max}} = 75$ @ (650, 250, 350, 350, 350)
- **Cuboid** $d = (173, 120, 200, 200, 200)$, $V_{\text{max}} = 60$ @ (850, 225, 650, 650, 650)
- **Cube** $d = (100, 100, 300, 300, 300)$, $V_{\text{max}} = 25$ @ (750, 750, 450, 450, 450)
- **Gaussian** $\sigma = (50, 150, 150, 150, 150)$, $V_{\text{max}} = 16$ @ (250, 250, 550, 550, 550)
- **Big Hypersphere** $r = 300$, $V_{\text{max}} = 12$ @ (350, 650, 650, 650, 650)
- **Background** $V = 0.1 + 4 \cdot 10^{-30} \cdot x_1^2x_2^2x_3^2x_4^2x_5^2$

1. $V=74.929@(655.00,253.72,347.83,348.57,349.59)$
2. $V=59.972@(850.04,224.99,650.00,649.99,654.56)$
3. $V=58.219@(849.97,225.01,587.08,650.01,650.02)$
4. $V=25.110@(750.00,749.99,450.00,450.01,450.01)$
5. $V=16.042@(245.45,253.44,552.51,542.58,544.75)$
7. ...
Plot Details

- **Parameters:** $x_1, \ldots, x_5 \in [0, 1000]$
- **Bins:** $50 \times 50$
- **PDF:** Breit-Wigner $\left( \frac{1}{1 + \Delta x_i^2 / \sigma^2} \right)$ with $\sigma = 100$
- **Number of Markov chains:** 9
- **Number of points per chain:** $10^7$
- **Number of function evaluations:** 33, 797, 153
- **Acceptance ratio:** 0.19
- **Final r (measure of convergence):** 1.815
- **CPU time (3 GHz):** 150 min