ENS DE LYON

Numerical frugality in optimization: Newton's methods in mixed precision

EUROPT 2026, Linz

G. Carrino, N. Brisebarre, T. Mary, E. Riccietti | Thursday 9th July, 2026

Joint work with

N. Brisebarre

T. Mary

E. Riccietti

Problem statement

Optimization problem

Solving the following optimization problem:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function.

A second-order optimizer

We want to solve this using **Newton's method**:

1. Compute $g_i = g(x_i)$

A second-order optimizer

We want to solve this using **Newton's method**:

1. Compute $g_i = g(x_i)$
2. Solve $H(x_i)d_i = -g_i$

A second-order optimizer

We want to solve this using **Newton's method**:

1. Compute $g_i = g(x_i)$
2. Solve $H(x_i)d_i = -g_i$
3. Update $x_{i+1} = x_i + d_i$

Newton's cost

Newton's method comes with two main sources of cost…

Solving the linear system

$$H(x_i)\, d_i = -g_i$$

Computing the Hessian

$$H(x_i)$$

Newton's cost

Newton's method comes with two main sources of cost…

Solving the linear system

$$H(x_i)\, d_i = -g_i$$

$$\downarrow$$

inexact Newton methods

$$\|H(x_i)d_i + g_i\| \leq \eta \|g_i\|$$

Computing the Hessian

$$H(x_i)$$

$$\downarrow$$

quasi-Newton methods

$$B(x) \approx H(x)$$

…but many possible approximations!

Quasi-Newton: an example

When solving least-squares problems,

Quasi-Newton: an example

When solving least-squares problems,

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|R(x)\|^2, \quad R : \mathbb{R}^n \to \mathbb{R}^m,$$

Quasi-Newton: an example

When solving least-squares problems,

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|R(x)\|^2, \quad R : \mathbb{R}^n \to \mathbb{R}^m,$$

the Hessian is given by

Quasi-Newton: an example

When solving least-squares problems,

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|R(x)\|^2, \quad R : \mathbb{R}^n \to \mathbb{R}^m,$$

the Hessian is given by

$$H(x) = J_R(x)^T J_R(x) + S(x), \quad S(x) = \sum_{i=0}^{m-1} R_i(x) \nabla^2 R_i(x)$$

Quasi-Newton: an example

When solving least-squares problems,

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|R(x)\|^2, \quad R : \mathbb{R}^n \to \mathbb{R}^m,$$

the Hessian is given by

$$H(x) = J_R(x)^T J_R(x) + \cancel{S(x)}, \quad S(x) = \sum_{i=0}^{m-1} R_i(x) \nabla^2 R_i(x)$$

Gauss-Newton discards this!

Reducing the cost

- ***Issue***: these variants can also be expensive, especially for large-scale problems.

Reducing the cost

- ***Issue***: these variants can also be expensive, especially for large-scale problems.
- ***Idea***: use **low precision floating-point arithmetic** to reduce the overall cost.

Floating-point precision

$$\mathsf{fl}(x \circ y) = (x \circ y)(1+\delta), \quad |\delta| \leq u, \quad \circ \in \{+, -, *, /\}$$

Floating-point arithmetic model.

Floating-point precision

$$\mathrm{fl}(x \circ y) = (x \circ y)(1+\delta), \quad |\delta| \le u, \quad \circ \in \{+, -, *, /\}$$

Floating-point arithmetic model.

Type	Size	Unit roundoff (u)	Approx. range
bfloat16	16 bits	4×10^{-3}	$10^{\pm 38}$
fp16	16 bits	5×10^{-4}	$10^{\pm 4}$
fp32	32 bits	6×10^{-8}	$10^{\pm 38}$
fp64	64 bits	1×10^{-16}	$10^{\pm 308}$

$\approx 30\times$

$\approx 4\times$

Floating point formats.

Reducing the cost

- ***Issue***: also these variants can be expensive, especially for large-scale problems.
- ***Idea***: use **low floating-point precision** to reduce the overall cost.
- ***Challenge***: how to do it without sacrificing convergence?

Reducing the cost

- ***Issue***: also these variants can be expensive, especially for large-scale problems.
- ***Idea***: use **low floating-point precision** to reduce the overall cost.
- ***Challenge***: how to do it without sacrificing convergence?

We can leverage **mixed precision**!

Mixed precision Newton's method

Error model

For each iteration i, the mixed precision Newton's step satisfies

$$\widehat{d}_i \coloneqq -\left(H(\widehat{x}_i) + E_i^H\right)^{-1}\left(g(\widehat{x}_i) + e_i^g\right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+ .$$

Error model

For each iteration i, the mixed precision Newton's step satisfies

$$\widehat{d}_i \coloneqq -\left(H(\widehat{x}_i) + E_i^H\right)^{-1}\left(g(\widehat{x}_i) + e_i^g\right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+.$$

- E_i^H: Hessian approximations and/or inexact linear solvers.

Error model

For each iteration i, the mixed precision Newton's step satisfies

$$\widehat{d}_i \coloneqq -\left(H(\widehat{x}_i) + E_i^H\right)^{-1}\left(g(\widehat{x}_i) + e_i^g\right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+ .$$

- E_i^H: Hessian approximations and/or inexact linear solvers.
- e_i^g: inexact gradient computations.

Error model

For each iteration i, the mixed precision Newton's step satisfies

$$\widehat{d}_i \coloneqq -\left(H(\widehat{x}_i) + E_i^H\right)^{-1}\left(g(\widehat{x}_i) + e_i^g\right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+.$$

- E_i^H: Hessian approximations and/or inexact linear solvers.
- e_i^g: inexact gradient computations.
- e_i^+: errors in the update step.

Local convergence analysis

$$\|\widehat{x}_{i+1} - x^*\| \leq \alpha_i \|\widehat{x}_i - x^*\|^2 + \beta_i \|\widehat{x}_i - x^*\| + \gamma_i,$$

Mixed precision Newton's method convergence rate.

Local convergence analysis

$$\|\widehat{x}_{i+1} - x^*\| \leq \boxed{\alpha_i} \|\widehat{x}_i - x^*\|^2 + \beta_i \|\widehat{x}_i - x^*\| + \gamma_i \,,$$

Mixed precision Newton's method convergence rate.

- α_i: standard Newton's quadratic convergence;

Local convergence analysis

$$\|\widehat{x}_{i+1} - x^*\| \leq \alpha_i \|\widehat{x}_i - x^*\|^2 + \beta_i \|\widehat{x}_i - x^*\| + \gamma_i,$$

Mixed precision Newton's method convergence rate.

- α_i: standard Newton's quadratic convergence;
- β_i: linear convergence influenced by Hessian errors;

Local convergence analysis

$$\|\widehat{x}_{i+1} - x^*\| \leq \alpha_i \|\widehat{x}_i - x^*\|^2 + \beta_i \|\widehat{x}_i - x^*\| + \gamma_i,$$

Mixed precision Newton's method convergence rate.

- α_i: standard Newton's quadratic convergence;
- β_i: linear convergence influenced by Hessian errors;
- γ_i: **limiting accuracy** mainly related to gradient errors.

Error model applications

Floating-point arithmetic Newton's method

We take into account a floating-point precision for each operation.

Floating-point arithmetic Newton's method

We take into account a floating-point precision for each operation.

1. Compute $g_i = g(x_i)$ in precision u_g

Floating-point arithmetic Newton's method

We take into account a floating-point precision for each operation.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Solve $H(x_i)d_i = -g_i$ in precision u_H

Floating-point arithmetic Newton's method

We take into account a floating-point precision for each operation.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Solve $H(x_i)d_i = -g_i$ in precision u_H
3. Update $x_{i+1} = x_i + d_i$ in precision u

Finite precision and error model

$$\widehat{d}_i \coloneqq -\Big(H(\widehat{x}_i) + E_i^H\Big)^{-1}\Big(g(\widehat{x}_i) + e_i^g\Big),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+ .$$

Mixed precision Newton's error model.

Finite precision and error model

$$\widehat{d}_i \coloneqq -\left(H(\widehat{x}_i) + E_i^H\right)^{-1}\left(g(\widehat{x}_i) + e_i^g\right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + \boxed{e_i^+}.$$

Mixed precision Newton's error model.

- $\boxed{\|e_i^+\|} \lesssim u$, by standard floating-point arithmetic model;

Finite precision and error model

$$\widehat{d}_i \coloneqq -\Big(H(\widehat{x}_i) + \boxed{E_i^H}\Big)^{-1}\Big(g(\widehat{x}_i) + e_i^g\Big),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + \boxed{e_i^+}.$$

Mixed precision Newton's error model.

- $\boxed{\|e_i^+\|} \lesssim u$, by standard floating-point arithmetic model;
- $\boxed{\|E_i^H\|} \approx O(u_H)$, if linear system is solved with a backward stable method;

Finite precision and error model

$$\widehat{d}_i \coloneqq -\left(H(\widehat{x}_i) + E_i^H\right)^{-1}\left(g(\widehat{x}_i) + e_i^g\right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+.$$

Mixed precision Newton's error model.

- $\|e_i^+\| \lesssim u$, by standard floating-point arithmetic model;
- $\|E_i^H\| \approx O(u_H)$, if linear system is solved with a backward stable method;
- $\|e_i^g\| \approx O(u_g)$.

Mixed precision setting

The setting of interest is

$$u_g \leq u \leq u_H$$

Mixed precision setting

The setting of interest is

$$u_g \leq \boxed{u} \leq u_H$$

- **Target precision**: the accuracy we aim to achieve.

Mixed precision setting

The setting of interest is

$$\boxed{u_g} \leq \boxed{u} \leq u_H$$

- **Target precision**: the accuracy we aim to achieve.
- **Accurate gradient** (u_g small): better limiting accuracy γ_i.

Mixed precision setting

The setting of interest is

$$\boxed{u_g} \leq \boxed{u} \leq \boxed{u_H}$$

- **Target precision**: the accuracy we aim to achieve.
- **Accurate gradient** (u_g small): better limiting accuracy γ_i.
- **Reduced precision linear solver** (u_H large): reduced cost per iteration.

Mixed precision setting

The setting of interest is

$$\boxed{u_g} \leq \boxed{u} \leq \boxed{u_H}$$

- **Target precision**: the accuracy we aim to achieve.
- **Accurate gradient** (u_g small): better limiting accuracy γ_i.
- **Reduced precision linear solver** (u_H large): reduced cost per iteration.
- **Trade-off**: potentially slower convergence.

Mixed precision setting

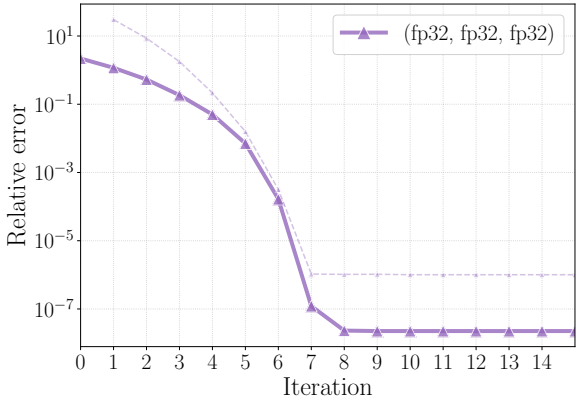
The setting of interest is

$$u_g \leq u \leq u_H$$

- **Target precision**: the accuracy we aim to achieve.
- **Accurate gradient** (u_g small): better limiting accuracy γ_i.
- **Reduced precision linear solver** (u_H large): reduced cost per iteration.
- **Trade-off**: potentially slower convergence.

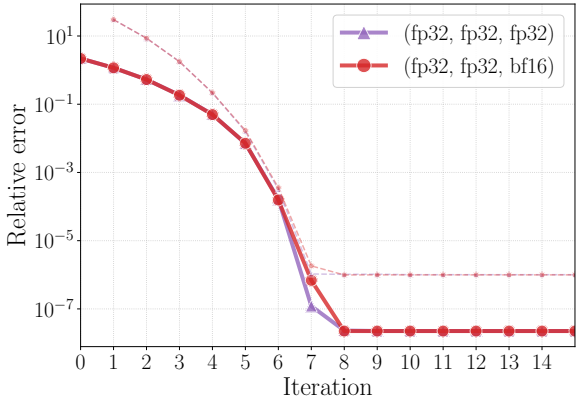
We will denote precision sets as (u_g, u, u_H)

Charts - Mixed precision Newton



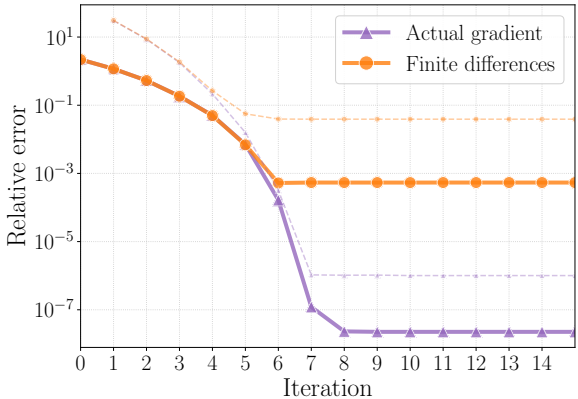
ENGVAL1. Mixed precision Newton relative error and its predicted rate.

Charts - Two mixed precision settings



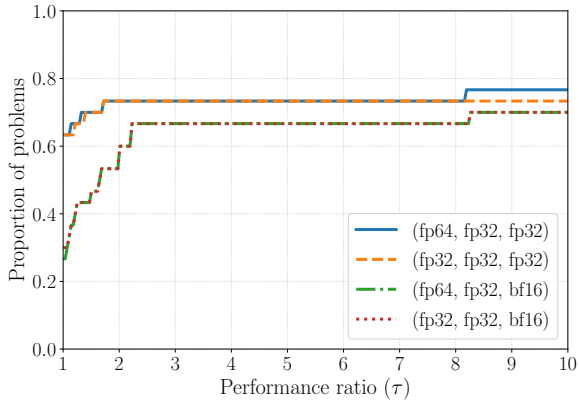
ENGVAL1. Different mixed precision settings relative error.

Charts - Finite differences



ENGVAL1. Finite differences vs true gradient relative error.

Charts - CUTEst performance profile



`CUTEst problems collection`. Performance profile of different mixed precision settings.

Error model extensions

Inexact Newton

$$\|H(x_i)d_i + g_i\| \leq \eta\|g_i\|$$

linear solver stopped early at tolerance η

Error model extensions

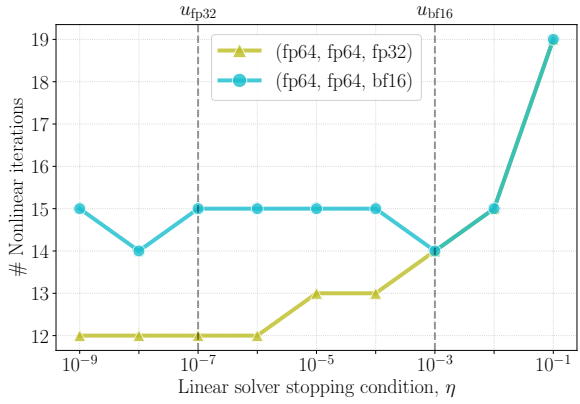
Inexact Newton

$$\|H(x_i)d_i + g_i\| \leq \eta \|g_i\|$$

linear solver stopped early at tolerance η

$$\|E_i^H\| \lesssim c_{\text{solver}} u_H + \frac{\|g(\widehat{x}_i)\|}{\|H(\widehat{x}_i)\|\|\widehat{d}_i\|}\eta$$

Charts - Inexact Newton



ENGVAL1. Inexact Newton: iterations vs. η.

Error model extensions

Gauss-Newton

$$H(x) = J_R(x)^T J_R(x) \cancel{+ S(x)}$$

Hessian approximation discarding $S(\widehat{x}_i)$

Error model extensions

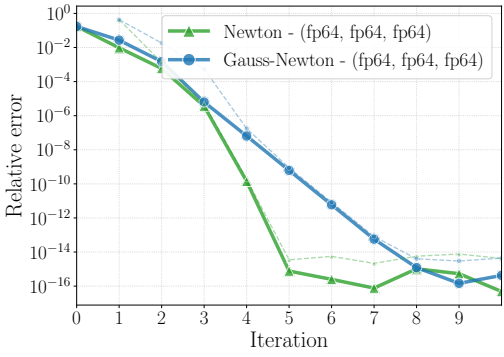
Gauss-Newton

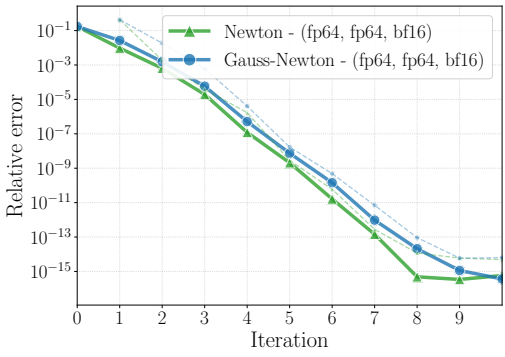
$$H(x) = J_R(x)^T J_R(x) \cancel{+ S(x)}$$

Hessian approximation discarding $S(\widehat{x}_i)$

$$\|E_i^H\| \leq c_{\text{solver}} u_H + \frac{\|S(\widehat{x}_i)\|}{\|H(\widehat{x}_i)\|}$$

Charts - Gauss-Newton





SINREG. Gauss-Newton vs. Newton's method.

Contributions

Our contribution

01 **Error model** — A comprehensive framework for mixed precision Newton's methods.

Our contribution

01 **Error model** — A comprehensive framework for mixed precision Newton's methods.

02 **Convergence analysis** — Rigorous study of the interplay between *precision* and *convergence rate*.

Our contribution

01 **Error model** — A comprehensive framework for mixed precision Newton's methods.

02 **Convergence analysis** — Rigorous study of the interplay between *precision* and *convergence rate*.

03 **Numerical validation** — Experiments confirming the practical benefits of mixed precision strategies.

Short-term directions

Scaling strategies

Scaled values allowing for lower precision linear solvers.

Short-term directions

Scaling strategies

Scaled values allowing for lower precision linear solvers.

HPC benchmarking

Practical implementations for large-scale optimization.

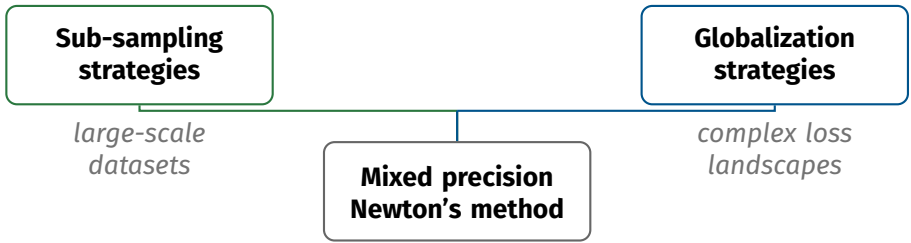
New perspectives

Towards machine learning

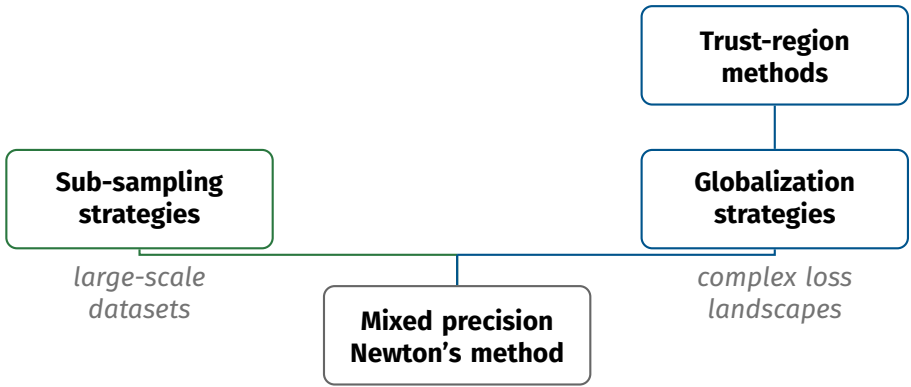
Mixed precision Newton's method

Towards machine learning

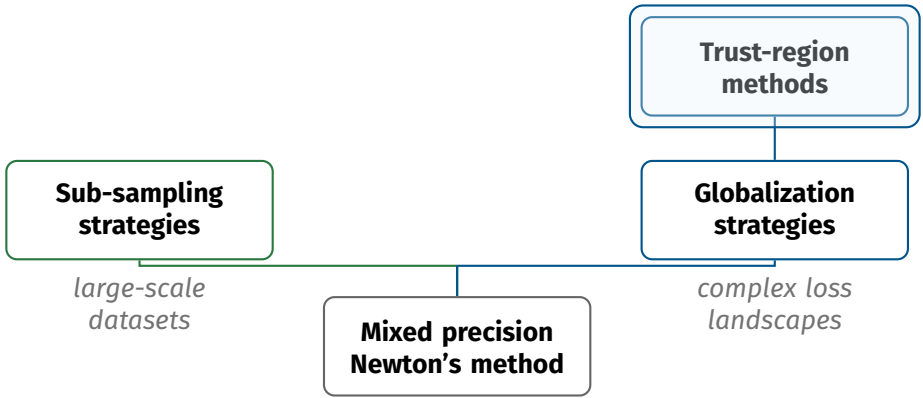
Towards machine learning



Towards machine learning



Towards machine learning



Trust-region methods

Model function $m_i(d) := f(\hat{x}_i) + g(\hat{x}_i)^T d + \frac{1}{2} d^T H(\hat{x}_i) d$

Trust-region methods

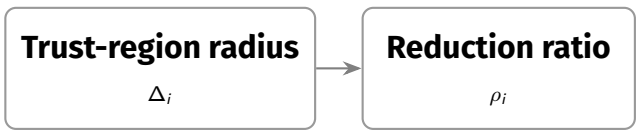
Model function $m_i(d) := f(\hat{x}_i) + g(\hat{x}_i)^T d + \frac{1}{2} d^T H(\hat{x}_i) d$

Trust-region radius

Δ_i

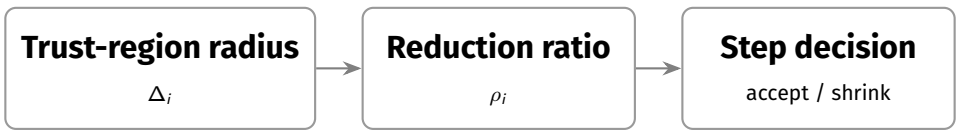
Trust-region methods

Model function $m_i(d) := f(\widehat{x}_i) + g(\widehat{x}_i)^T d + \frac{1}{2} d^T H(\widehat{x}_i) d$



Trust-region methods

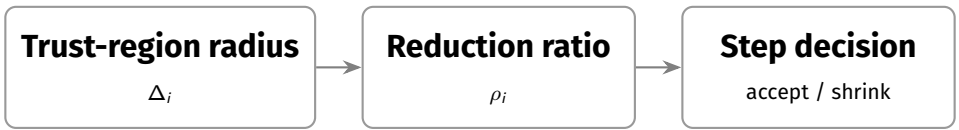
Model function $m_i(d) := f(\widehat{x}_i) + g(\widehat{x}_i)^T d + \frac{1}{2} d^T H(\widehat{x}_i) d$



$\rho_i = \dfrac{f(\widehat{x}_i) - f(\widehat{x}_i + \widehat{d}_i)}{m_i(0) - m_i(\widehat{d}_i)}$ compares actual vs. predicted decrease.

Trust-region methods

Model function $m_i(d) := f(\widehat{x}_i) + g(\widehat{x}_i)^T d + \frac{1}{2} d^T H(\widehat{x}_i) d$



$\rho_i = \dfrac{f(\widehat{x}_i) - f(\widehat{x}_i + \widehat{d}_i)}{m_i(0) - m_i(\widehat{d}_i)}$ compares actual vs. predicted decrease.

Accept the step if ρ_i is large; otherwise shrink Δ_i and recompute.

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

1. Compute $g_i = g(x_i)$ in precision u_g

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Minimize $m_i(d_i)$ s.t. $\|d_i\| \leq \Delta_i$ in precision u_H

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Minimize $m_i(d_i)$ s.t. $\|d_i\| \leq \Delta_i$ in precision u_H
3. Compute ρ_i in precision u_f

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Minimize $m_i(d_i)$ s.t. $\|d_i\| \leq \Delta_i$ in precision u_H
3. Compute ρ_i in precision u_f
4. Update Δ_i based on ρ_i

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Minimize $m_i(d_i)$ s.t. $\|d_i\| \leq \Delta_i$ in precision u_H
3. Compute ρ_i in precision u_f
4. Update Δ_i based on ρ_i
5. Update $x_{i+1} = x_i + d_i$ based on ρ_i in precision u

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Minimize $m_i(d_i)$ s.t. $\|d_i\| \leq \Delta_i$ in precision u_H
3. Compute ρ_i in precision u_f
4. Update Δ_i based on ρ_i
5. Update $x_{i+1} = x_i + d_i$ based on ρ_i in precision u

Mixed precision trust-region algorithm

We modify the previous algorithm including a trust-region approach.

1. Compute $g_i = g(x_i)$ in precision u_g
2. Minimize $m_i(d_i)$ s.t. $\|d_i\| \leq \Delta_i$ in precision u_H
3. Compute ρ_i in precision u_f
4. Update Δ_i based on ρ_i
5. Update $x_{i+1} = x_i + d_i$ based on ρ_i in precision u

Application: logistic regression

Binary classification task via **logistic regression**.

Application: logistic regression

Binary classification task via **logistic regression**.

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{m} \sum_{j=1}^{m} \log\Big(1 + \exp(-b_j \, a_j^\top x)\Big) + \frac{\lambda}{2} \|x\|^2$$

Application: logistic regression

Binary classification task via **logistic regression**.

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{m} \sum_{j=1}^{m} \log\Big(1 + \exp(-b_j \, a_j^\top x)\Big) + \frac{\lambda}{2} \|x\|^2$$

- $(a_j, b_j)_{j=1}^m$: feature vectors and binary labels ($b_j \in \{-1, 1\}$);
- λ: regularization parameter (set to 10^{-3}).

Application: logistic regression

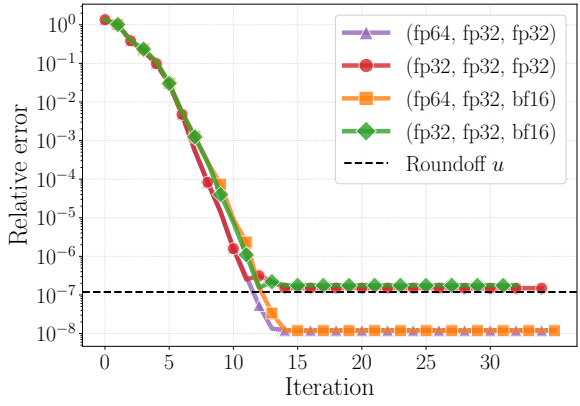
Binary classification task via **logistic regression**.

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{m} \sum_{j=1}^{m} \log\Big(1 + \exp(-b_j \, a_j^\top x)\Big) + \frac{\lambda}{2} \|x\|^2$$

- $(a_j, b_j)_{j=1}^m$: feature vectors and binary labels ($b_j \in \{-1, 1\}$);
- λ: regularization parameter (set to 10^{-3}).

Australian Credit Approval dataset (UCI/Statlog)
690 instances, 14 attributes

Preliminary experiments



`Australian Dataset`. Mixed precision TR-Newton relative error.

Thank you!

Questions?

Gradient norm decrease

$$\|g(\widehat{x}_{i+1})\| \leq \phi_i \|g(\widehat{x}_i)\| + \psi_i.$$

Mixed precision Newton's method gradient norm decay.

- ϕ_i: "quadratic"/linear convergence influenced by Hessian errors;
- $\psi_i \approx \epsilon_i^g + \epsilon_i \|H(\widehat{x}_i)\| \|\widehat{x}_i\| \rightarrow$ **Potential stopping criterion!**

Convergence theorem assumptions

$$\|e_i^H\| \leq \epsilon_i^H \|H(\widehat{x}_i)\|, \quad \|e_i^g\| \leq \epsilon_i^g, \quad \|e_i^+\| \leq \epsilon_i(\|\widehat{d}_i\| + \|\widehat{x}_i\|).$$

At a given iteration

- $\epsilon_i^H \kappa(H(\widehat{x}_i)) < 1$

General local convergence

- $\epsilon_i^H \kappa(H(\widehat{x}_i)) < 1 \quad \forall i$
- $\theta_i := \alpha_i \|\widehat{x}_i - x^*\| + \beta_i < 1 \quad \forall i$
- until $\|\widehat{x}_i - x^*\| \leq \frac{\gamma_i}{1-\theta_i}$.

Numerical experiments problems

$$f(x) = 3 + \sum_{k=0}^{n-2} (x_k^2 + x_{k+1}^2)^2 - 4x_k.$$

ENGVAL1 function (from CUTEst dataset).

$$R(x) = x_0 z + \sum_{k=1}^{\lceil \frac{n-1}{2} \rceil} x_{2k-1} z^{k+1} + \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} x_{2k} \sin(x_{2k} z),$$

SINREG least squares residual, with z fixed data.

Gradient descent - Error model

Question: 2nd-order vs 1st-order methods in mixed precision?

Gradient descent - Error model

Question: 2nd-order vs 1st-order methods in mixed precision?

For each iteration i, the mixed precision GD's step satisfies

$$\widehat{d}_i := -\tau_i \left(g(\widehat{x}_i) + e_i^g \right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+ .$$

- e_i^g: inexact gradient computations.

Gradient descent - Error model

Question: 2nd-order vs 1st-order methods in mixed precision?

For each iteration i, the mixed precision GD's step satisfies

$$\widehat{d}_i := -\tau_i \left(g(\widehat{x}_i) + e_i^g \right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+.$$

- e_i^g: inexact gradient computations.
- e_i^+: errors in the update step.

Gradient descent - Error model

Question: 2nd-order vs 1st-order methods in mixed precision?

For each iteration i, the mixed precision GD's step satisfies

$$\widehat{d}_i := -\tau_i\left(g(\widehat{x}_i) + e_i^g\right),$$
$$\widehat{x}_{i+1} = \widehat{x}_i + \widehat{d}_i + e_i^+.$$

- e_i^g: inexact gradient computations.
- e_i^+: errors in the update step.
- τ_i: step size.

Convergence analysis of GD

Assuming f L-smooth, μ-strongly convex, and $\tau_i \in (0, 1/3L)$, we have

$$\|\widehat{x}_{i+1} - x^*\|^2 \leq \alpha_i^{GD} \|\widehat{x}_i - x^*\|^2 + \beta_i^{GD} \|\widehat{x}_i - x^*\| + \gamma_i^{GD},$$

Mixed precision gradient descent convergence rate.

Convergence analysis of GD

Assuming f L-smooth, μ-strongly convex, and $\tau_i \in (0, 1/3L)$, we have

$$\|\widehat{x}_{i+1} - x^*\|^2 \leq \boxed{\alpha_i^{GD}} \|\widehat{x}_i - x^*\|^2 + \beta_i^{GD} \|\widehat{x}_i - x^*\| + \gamma_i^{GD},$$

Mixed precision gradient descent convergence rate.

- α_i^{GD}: linear convergence term;

Convergence analysis of GD

Assuming f L-smooth, μ-strongly convex, and $\tau_i \in (0, 1/3L)$, we have

$$\|\widehat{x}_{i+1} - x^*\|^2 \leq \boxed{\alpha_i^{GD}}\|\widehat{x}_i - x^*\|^2 + \boxed{\beta_i^{GD}}\|\widehat{x}_i - x^*\| + \gamma_i^{GD},$$

Mixed precision gradient descent convergence rate.

- α_i^{GD}: linear convergence term;
- β_i^{GD}: sub-linear convergence influenced by gradient errors;

Convergence analysis of GD

Assuming f L-smooth, μ-strongly convex, and $\tau_i \in (0, 1/3L)$, we have

$$\|\widehat{x}_{i+1} - x^*\|^2 \leq \boxed{\alpha_i^{GD}}\|\widehat{x}_i - x^*\|^2 + \boxed{\beta_i^{GD}}\|\widehat{x}_i - x^*\| + \boxed{\gamma_i^{GD}},$$

Mixed precision gradient descent convergence rate.

- α_i^{GD}: linear convergence term;
- β_i^{GD}: sub-linear convergence influenced by gradient errors;
- γ_i^{GD}: **limiting accuracy**.

Convergence analysis of GD

Assuming f L-smooth, μ-strongly convex, and $\tau_i \in (0, 1/3L)$, we have

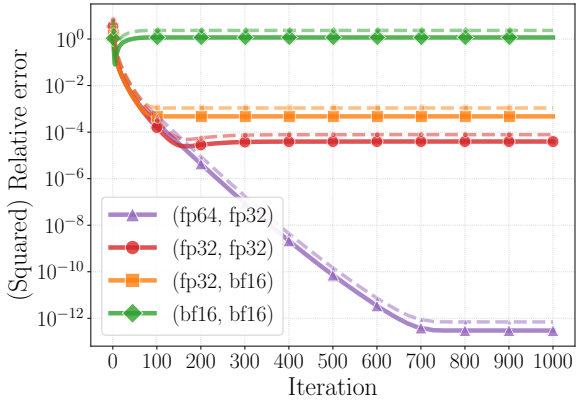
$$\|\widehat{x}_{i+1} - x^*\|^2 \leq \boxed{\alpha_i^{GD}}\|\widehat{x}_i - x^*\|^2 + \boxed{\beta_i^{GD}}\|\widehat{x}_i - x^*\| + \boxed{\gamma_i^{GD}},$$

Mixed precision gradient descent convergence rate.

- α_i^{GD}: linear convergence term;
- β_i^{GD}: sub-linear convergence influenced by gradient errors;
- γ_i^{GD}: **limiting accuracy**.

Observation: $\lim_{i\to\infty} \tau_i = 0 \implies \lim_{i\to\infty} \gamma_i^{GD} \approx u^2\|x^*\|^2$.

Preliminary experiments



`Quadratic function`. Mixed precision GD using finite differences and decaying τ_i.