Algebraic Identities for '' Accelerated Objective Gap and Gradient NormConvergence for Gradient Descent via Long Steps '' [Grimmer, Shu, Wang 2024]

Section 3

In[ ]:=
```
(* 3.1 *)
Remove["Global`*"];
ClearAll;
rho = 1 + Sqrt[2];
beta[k_] := 1 + rho^(k - 1);
Simplify[(rho^k - 1) * (1 - 1 / rho) - beta[k] / rho - (rho^(k - 1) - 1)]
Simplify[(rho^k - 1) * (1 - rho) + (rho^(k + 1) - 1) / rho + rho^(k - 1) - 1]
```

Out[ ]= 0

Out[ ]= 0

In[ ]:=
```
(* 3.2 *) (* Common setup definitions *)
Remove["Global`*"];
ClearAll;
rho = 1 + Sqrt[2];
beta[k_] := 1 + rho^(k - 1);
x[n] = {1, 0, 0, 0};
g[n] = {0, 1, 0, 0};
x[2 n + 1] = {0, 0, 1, 0};
g[2 n + 1] = {0, 0, 0, 1};
H[a_, b_] := (KroneckerProduct[a, b] + KroneckerProduct[b, a]) / 2;

(*Additional quantities depending on step lengths*)
x[n + 1] = x[n] - beta[k] * g[n];

(* coefficients on f_n and f_{2n+1} in the error term*)
Simplify[{(rho^k - 1) (1 - 1 / rho) - beta[k] / rho - (rho^(k - 1) - 1),
  (rho^k - 1) (1 - rho) + 1 / rho * (rho^(k + 1) - 1) + rho^(k - 1) - 1}]

(* quadratic form in the error term *)
Thetaerror = (1 / (2 rho)) * H[x[2 n + 1], x[2 n + 1]] +
   (rho^k (rho^(k + 1) - 1)) / 2 * H[g[2 n + 1], g[2 n + 1]] -
   beta[k] / rho (H[g[n], g[n]] / 2 + H[g[n], -x[n]]) - (1 / (2 rho)) H[x[n], x[n]] -
   rho^(k - 1) (rho^k - 1) / 2 H[g[n], g[n]] - rho H[x[n + 1] - x[2 n + 1], -x[n + 1]] -
   rho / 2 H[x[2 n + 1] - x[n + 1], x[2 n + 1] - x[n + 1]] -
   (rho^(k + 1) (rho^k - 1)) / 2 * H[g[2 n + 1], g[2 n + 1]];
Xiterm = H[x[2 n + 1] - x[n + 1], x[n] + x[2 n + 1]] - (rho^k - 1) / 2 H[g[n], g[n]] -
   (rho^k - 1) / 2 H[g[2 n + 1], g[2 n + 1]] + H[x[n + 1] - x[2 n + 1], g[n] + g[2 n + 1]];
Deltaterm = H[x[2 n + 1] - x[n], g[2 n + 1] - rho^(k - 1) g[n]] -
   beta[k] / 2 × H[g[2 n + 1] - g[n], g[2 n + 1] - g[n]];
FullSimplify[MatrixForm[Thetaerror + Xiterm + Deltaterm]]
```

Out[ ]= {0, 0}

Out[ ]//MatrixForm=

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Section 4

In[ ]:=
```
(* 4.1 *)
Remove["Global`*"];
ClearAll;
rho = 1 + Sqrt[2];
beta[k_] := 1 + rho^(k - 1);
x[n] = {1, 0, 0, 0};
g[n] = {0, 1, 0, 0};
x[2 n + 1] = {0, 0, 1, 0};
g[2 n + 1] = {0, 0, 0, 1};
H[a_, b_] := (KroneckerProduct[a, b] + KroneckerProduct[b, a]) / 2;
(*Additional quantities depending on step lengths*)
x[n + 1] = x[n] - alpha[k] * g[n];
r[k] = 2 (alpha[k] - 1)^2 / (beta[k + 1] - alpha[k]);
r[k + 1] = 2 (beta[k + 1] - 1)^2 / (beta[k + 1] - alpha[k]);
gamma = r[k + 1] / rho^(2 * k - 1);
(*Quadratic form in the error term*)
ErrorTheta = - (r[k] - 1) / 2 * H[g[n], g[n]] + (r[k + 1] - 1) / 2 * H[g[2 n + 1], g[2 n + 1]] +
   gamma / rho * H[x[n + 1] - x[2 n + 1], x[n + 1]] -
   gamma / (2 * rho) * H[x[2 n + 1] - x[n + 1], x[2 n + 1] - x[n + 1]] -
   gamma * rho^(k - 1) * (rho^k - 1) / 2 * H[g[2 n + 1], g[2 n + 1]];
ErrorXi =
  gamma / (2 * rho) * (H[x[2 n + 1] - x[n + 1], x[n] + x[2 n + 1]] - (rho^k - 1) / 2 * H[g[n], g[n]] -
     (rho^k - 1) / 2 * H[g[2 n + 1], g[2 n + 1]] + H[x[n + 1] - x[2 n + 1], g[n] + g[2 n + 1]]);
ErrorDelta = gamma / (2 * rho) * (H[g[2 n + 1], x[2 n + 1] - x[n]] -
      1 / 2 * H[g[2 n + 1] - g[n], g[2 n + 1] - g[n]]) - (gamma * rho^(k - 1) / 2 - 1) *
    (H[g[n], x[2 n + 1] - x[n]] + 1 / 2 * H[g[2 n + 1] - g[n], g[2 n + 1] - g[n]]);
FullSimplify[MatrixForm[ErrorTheta + ErrorXi + ErrorDelta],
 {beta[k] < alpha[k] < beta[k + 1], k > 0}]
```

Out[ ]//MatrixForm=

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Section 5

In[ ]:=
```
(* 5.1 *)
(* Common setup definitions *)
Remove["Global`*"];
ClearAll;
rho = 1 + Sqrt[2];
beta[k_] := 1 + rho^(k - 1);
x[n] = {1, 0, 0, 0, 0};
g[n] = {0, 1, 0, 0, 0};
x[2 n + 1] = {0, 0, 1, 0, 0};
g[2 n + 1] = {0, 0, 0, 1, 0};
diff = {0, 0, 0, 0, 1}; (* this is y[2n+1] - y[n+1] *)
H[a_, b_] := (KroneckerProduct[a, b] + KroneckerProduct[b, a]) / 2;
(*Additional quantities depending on step lengths*)
x[n + 1] = x[n] - alpha[k] * g[n];
y[n] = 1 / Sqrt[r[k + 1]] * x[n];
y[n + 1] = y[n] - beta[k + 1] / Sqrt[r[k + 1]] * g[n];
y[2 n + 2] = y[n + 1] + diff;
r[k] = 2 (alpha[k] - 1)^2 / (beta[k + 1] - alpha[k]);
r[k + 1] =  2 (beta[k + 1] - 1)^2 / (beta[k + 1] - alpha[k]);

(*Quadratic form in the error term*)Error = 1 / 2 * H[y[2 n + 2], y[2 n + 2]] -
   (beta[k + 1] / r[k + 1]) * (H[g[n], g[n]] / 2 + H[g[n], -x[n]]) -
   1 / (2 r[k + 1]) H[x[n], x[n]] - rho^k (rho^k - 1) / (2 * r[k + 1]) H[g[n], g[n]] +
   1 / Sqrt[r[k]] * H[y[n + 1] - y[2 n + 2], x[n + 1]] -
   1 / 2 * H[y[n + 1] - y[2 n + 2], y[n + 1] - y[2 n + 2]] +
   (1 / Sqrt[r[k]] - 1 / Sqrt[r[k + 1]]) H[y[2 n + 2] - y[n + 1], x[n] - g[n]] -
   (1 - Sqrt[r[k]] / Sqrt[r[k + 1]]) H[g[n], g[n]] / 2;
FullSimplify[MatrixForm[Error], {beta[k] < alpha[k] < beta[k + 1], k > 0}]
```

Out[ ]//MatrixForm=

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$